The cognitive mechanisms of normal and pathological forgetting

Andreea Stamate

Doctor of Philosophy

The University of Edinburgh

2020

DEDICATION

This thesis is dedicated to my mum, who always believed in me more than what is reasonably expected.

Thank you for instilling in me the value of curiosity and for always doing everything possible to help me fulfil my dreams.

ACKNOWLEDGEMENTS

I would like to thank several amazing people that I was lucky enough to meet along this academic voyage (I hope I am not *forgetting* too many of you). I am in great debt to all the physicians for their help with the recruitment of patients, as well as the other key individuals which gave me access to the other populations included in these studies. Equally important are the hundreds of people who took the time to participate in the experiments presented in the following chapters. My thesis would have been considerably shorter without you.

As Mr. Newton once said, "If I have seen any further [in life], 'tis by standing on the shoulders of giants”. I have been extremely fortunate to receive support from some inspiring researchers. I am extremely grateful to Professor Alan Baddeley and to my supervisors, Professors Maria Brandimonte, Robert Logie and Sergio Della Sala, who have mentored me and showed me how to conduct good science. All your ideas and feedback significantly improved the quality of this thesis, but more importantly it helped me become a better researcher. I would like to give special thanks to my main supervisor Professor Sergio Della Sala who was kind and patient enough, not only to initially take on a master student with no previous knowledge of cognitive neuroscience, but subsequently encourage and support me in becoming a PhD student. Thank you for opening my mind to better horizons.

Finally, I am very grateful for my boyfriend, family and friends, who have kept me sane and gracefully supported me throughout this process.

Contents

LAY SUMMARY

The process of remembering information over long-term involves the conversion of just learned information (working memory) into long-term memory. These working memories are converted into long-term ones through a consolidation process. Recent evidence suggests that this consolidation process starts with the encoding of an event and may last up to months or even years. Emerging evidence further proposes that this process is circular, not linear. Specifically, each time we bring a consolidated memory to mind, that memory returns in a modifiable state and then goes through a re-consolidation process. During this re-consolidation the original memory can be affected by exposure to new information, thus the original memory can either be strengthened or weakened when we bring it to mind. While the act of repeated retrieval is key to forming strong memories, in some instances it can have a harmful effect. For example, when we bring to mind (retrieve, or test) only a part of a memory, such as when we only talk about certain details of a previous event, this can (in some circumstances) cause forgetting of the other details of that event.

The first three experiments in my thesis were designed to investigate why at times this selective retrieval causes forgetting while at others it strengthens memory. Knowledge about this topic can be of interest to several disciplines. For example, in education, where it can indicate more effective ways to study or, in clinical settings, whereby repeated tests are needed for medical diagnosis and follow-up. The finding from this kind of research that even testing only a sample of information can improve memory might be seen as an advantage from an educational perspective. This could, however, prove problematic in clinical assessments as it would mask any decline in underlying cognitive abilities across repeated assessments. Therefore, we must understand the conditions when sample testing will enhance and when it will reduce memory performance.

I showed that repeatedly retrieving sub-parts of prose material over the course of 1-month can reduce forgetting for both younger and older adults. There were two conditions in the experiments, half the participants were tested after one day, one week and one month, on different sections of the material tested on each occasion. The other half were tested only 1-month later. Experiment 1 used coherent material (four short stories). Experiment 2 used material with a disrupted narrative (16 sentences) using the same two conditions. All groups showed significant forgetting over one month, however, groups tested repeatedly showed less forgetting (irrespective of age or material). When retested only at 1-month, older adults showed an accelerated rate of forgetting compared to younger adults.

In a third experiment, I investigated whether this type of testing would also enhance long-term memory performance in Alzheimer’s Disease (AD) and whether people with AD will show an accelerated rate of forgetting compared to healthy adults of the same age. The same design as Experiments 1 and 2 was used, together with a simplified version of the material used in Experiment 1. The use of the two conditions (repeated testing vs. a single retest at one month) allowed me to disentangle the repetition effects from the actual forgetting rates. AD patients took significantly longer to learn the material, which showed that they have learning deficits. The AD groups, after having learned the material to a similar level as the healthy participants, forgot the information at the same rate (irrespective of the testing condition). Additionally, when the material was tested repeatedly this improved their memory at 1-month (the condition with retrieval practice). These results suggest that repeated selective retrieval of memories also benefits AD patients.

In these experiments, I had to ask some participants to study the material more than other participants (e.g., AD participants took longer to learn to the same level as healthy older adults). In the next two experiments, I investigated if these additional learning opportunities would influence subsequent forgetting rates.

The 4th experiment successfully replicated a classic study on this topic. The study compares participants who are given additional learning opportunities (three) to participants who are only given one. Participants are then tested on one of four different testing intervals (immediately after learning, 1-day, 5-days, or 10-days). I successfully replicated this study and found that varying the number of learning opportunities does not influence forgetting rates. Given recent concerns about the reliability of findings from single research studies, replications of studies are a basic requirement for scientific integrity by showing that a previously reported finding can be reproduced.

In the 5th Experiment, I investigated this topic further, taking into consideration individual differences in learning capacity across participants. For example, we know that people have different learning capacities, while some may only study once for a test and get a high mark (e.g., 9 out of 10), others will not perform as well (e.g., only score a 6 after having studied the material once). In the present experiment the maximum number participants could have scored on the memory test was 13. Hence, participants were grouped according to their learning capacity (higher learners, i.e., 9 and 10 /13, vs. lower learners i.e., 6 and 7 /13). After having been grouped, half of the participants were given more learning opportunities (2 exposures), while the other half was not. The results showed that varying the number of learning opportunities does influence the forgetting rate. Thus, more learning opportunities lead to less forgetting, and fewer learning opportunities lead to more forgetting. This was only the case for higher learners, as the rate of forgetting remained similar for lower learners. Slower learners reached a similar initial score (on the initial test) as that of the fast learners when given more learning opportunities, however, when tested again after 1-month the score was still higher for the faster learners.

Memory researchers make the distinction between types of memories. We do not only forget information from the past we can also forget ‘future information’. That is, we can either forget what we have done or forget what we have to do. The latter is called prospective memory, or memory for future intentions. Though the two types of memory share many similarities, a major distinction between them, and the main source of forgetting in prospective memory, pertains to how we bring the information to memory. While past memories are usually brought to mind by an external request to remember (specifically when this is done in the laboratory), future intentions seldom are given an external ‘prompt’.

Experiment 6 investigated whether the retrieval of an intended prospective memory action would be facilitated if participants repeated back the prospective memory task multiple times at encoding (i.e., learned the prospective memory task more). During the initial test session participants were tested on the sentence material (used in Experiment

2) and given instructions for the prospective memory task. For the prospective memory task participants had to remember to deliver a message ('The sun is shining'). To test whether the repetition of the task would improve the retrieval of the intended action, half the participants were asked to repeat the message ('The sun is shining', out-loud) three times. The remaining participants only repeated the message back once. All participants were told they would receive a phone call at a later time to test their memory for the sentences but that they should also remember to deliver the prospective memory message when receiving the call. Participants were unaware when they will receive the phone call (1 month later) or that the prospective memory test was the main purpose of the experiment. The results showed that repeatedly bringing to mind (repeated retrieval) the prospective memory intention during its formation stage (encoding) improved participants ability to retrieve the intended prospective memory action (message) 1-month later.

The final experiment (Experiment 7) assessed younger and older adults' memory for future intentions in the context whereby these intentions are routinely performed (e.g., such as taking medication every day). We know that in this type of memory the main difficulty lies in remembering whether an action has already been performed (did I take the medication already, or was I just thinking of taking it) or not. Due to this difficulty, older adults tend to repeat the action (take medication again), while younger adults are more likely to simply forget to perform the action. In this last experiment, I aimed to enhance participants' memory of the just performed action (in this case a mouse click) by giving them additional contextual information (a different animal picture appeared for each click) in an attempt to reduce repetition errors or omission errors. All participants took part in an experiment formed of 12 trials, 2 minutes each, they had to remember to press the mouse button (only once) on each of the trials (the high number of trials is used to mimic a real-life routine). On half of the trials, when they clicked, the words 'you clicked' together with a different figure appeared (to make these trials more memorable) aiming to make them remember they had performed the action. The results of this experiment showed that making an action more memorable during its execution can increase younger and older adults' memory performance and thus reduce the likelihood of them committing errors (forgetting to perform the action or repeating it).

ABSTRACT

The objectives of this PhD were to investigate forgetting, how to measure forgetting and what are the underlying mechanisms behind this effect.

1. Following the progress of forgetting over time will require repeated testing. There is strong evidence that this may enhance recall, while opposing evidence, from part-set cueing studies suggests that probing one item may reduce the memorability of others within the set. It appears that whenever we test memory, we change it, but how? Experiment 1 and Experiment 2 compare long-term forgetting in healthy younger and older individuals and investigates whether repeated partial testing will enhance long-term memory performance based on the level of semantic coherence or integration of material to be remembered, possibly via a relearning or a priming effect.

2. Some previous studies showed that, under particular experimental conditions, patients suffering from a range of memory deficits (i.e., amnesia) appear to retain what they have learnt as well as controls (Huppert, & Piercy, 1978; 1979; Squire, 1981; Kopelman, 1985; Freed, Corkin & Cohen, 1987; Frisk & Milner, 1990a; 1991; Greene et al., 1995). A challenge to this view comes from a subsample of epileptic patients, who have been found to show accelerated long-term forgetting (ALF), in some cases showing normal learning over a period of 30 minutes to one hour, followed by dramatic loss of information later on (Butler & Zeman, 2008). Whether ALF applies to other patient populations remains to be established. Experiment 3 examines long-term forgetting as well as the effect of repeated partial testing in healthy people and in people with Alzheimer’s Disease (AD). Assessing whether AD patients present with ALF and whether repeated partial testing delays long-term forgetting in both groups.

3. A frequent assumption is that when comparing two groups they should be matched on level of initial learning. At the same time, several influential studies from the early

literature suggest that initial degrees of learning do not influence the rate of forgetting in the long-term (e.g., Slamecka & McElree, 1983; Slamecka, 1985). Experiment 4 is a replication of Experiment 1 from Slamecka & McElree's (1983) classic study.

4. Achieving a particular degree of learning at encoding will require more exposures for lower performing individuals compared to high performing individuals. Therefore, Experiment 5 aimed to examine whether varying the degrees of learning at encoding will differentially affect participants long-term forgetting rate based on their individual learning capacity.

5. An important classification of long-term memory is based on its temporal direction. While retrospective memory (RM) deals with past information, the remembering of future intentions depends on prospective memory (PM). The similarities between these two memory types have been frequently raised, yet an important distinction lies in their evaluation. When we measure RM tasks (in the laboratory) the participant is specifically directed to retrieve information, while when measuring PM tasks, the retrieval of the intended action is self-initiated. Experiment 6 investigated whether the retrieval of an intended PM action would be facilitated if participants repeated back the PM task multiple times at encoding (i.e., learned the PM task more).

6. In everyday life we typically only remember the gist of events that are encountered once and processed incidentally. Similarly, even events that are encountered repeatedly can be processed incidentally, with much of the rich contextual details being forgotten. In the forgetting literature, one class of such repetitive events such as taking medication daily, refer to habitual PM. In habitual PM the necessity of initiating (or not) a certain action is highly dependent on the accurate memory of the previously performed action (Marsh et al., 2007; McDaniel et al., 2009). The aim of Experiment 7 was to examine forgetting of habitual PM tasks and its underlying mechanisms, as well as to devise a method to enhance the memory of previously performed actions in habitual PM.

CHAPTER 1: REVIEW OF THE LITERATURE

1.1. Defining forgetting.

There has been a long search for a general description of forgetting. The study of the processes between memory retention and the delay between study and test is arguably one of psychology's oldest questions (Ebbinghaus, 1885/1964), with studies focusing on different forms of memory, exploring different types of materials and modalities.

One of the first researchers to study forgetting was Ebbinghaus (1885). He plotted forgetting curves of his own remembering over several days and found that forgetting occurs in a systematic manner, beginning rapidly and then levelling off. Subsequent research on forgetting, looking at either healthy participants or patients, and assessing either 'short-term' or 'long-term' forgetting, typically showed a similar pattern (e.g., Baddeley & Warrington, 1970; Butters & Cermak, 1980; Butters & Cermak, 1986; Kopelman & Stanhope, 1997).

Tulving (1974) defined forgetting as 'the inability to recall something now that could be recalled on an earlier occasion' (p. 74). This definition, logically implies, that only information which has been learned in the first place can be considered to be forgotten (Roediger, Weinstein, & Agarwal, 2010). When applying Tulving's (1974) definition in the context of a common memory paradigm, composed of an encoding or learning phase, a retention interval, and a final retrieval test, forgetting can be operationalised as a lower level of performance on a test compared to the level of performance on an earlier test. In an attempt to answer why and how forgetting occurs, researchers have manipulated various aspects of this paradigm. They did so either by manipulating the way information is encoded, the length between testing intervals, what happens during these intervals, or the nature of the final test (Storm, 2018).

An important distinction in the definition of forgetting, specifically from a theoretical perspective, has to be made between availability and accessibility of stored information

(Tulving & Pearlstone, 1966). This distinction indicates instances when memory could either be temporarily inaccessible (when a memory is stored but retrieval cannot be initiated) or physically unavailable (that is, memory is lost). The vast number of events, conversations, and information encountered in everyday life makes it unlikely that it could all be lying dormant in memory, waiting for the right cue to activate it, thus forgetting as a consequence of complete loss of certain information probably exists (Roediger, Weinstein, & Agarwal, 2010). Nonetheless, experimental evidence showing forgetting in a permanent state, as in physically unavailable, is difficult to provide in a way that can rule out alternative accounts, such as that information could be later recalled if given a different cue or in a different condition. That is why much research has focused on studying forgetting in terms of an accessibility impairment.

1.1.1. Forgetting as retrieval failure.

A popular idea in the older literature maintained that forgetting only reflects temporary inaccessibility, due to a retrieval failure, and that all encoded and stored information, in a way, persists within the nervous system. While this idea is viewed less favourably by current research, at the time, Tulving and Pearlstone (1966) presented a convincing experiment providing evidence in its favour. When testing a group of participants on a free recall word task, Tulving and Pearlstone (1966) found that participants forgot (failed to retrieve) about 29 words (19 words recalled out of 48 words in total). The authors tested the possibility that traces of the un-retrieved words were still stored but could not be retrieved with the minimal cues of free recall, by giving them stronger cues (category names). After these cues were given, participants were able to recall a mean of 35.9 words (12.1 words forgotten). Thus, in this second condition, participants recalled nearly twice as many words as in the first condition, providing evidence that (at least) some of the forgetting resulted from a temporary inaccessibility. Such examples of ‘reversing’ forgetting were likely to support the views in the late 1970s that forgetting was mostly due to retrieval failures (Roediger, Weinstein & Agarwal, 2010). A further point that can be made with this example is by considering the 25% of the words, which were still not retrieved. While additional methodological means (e.g., implicit tests) could have been employed to generate retrieval of more of the un-retrieved words, thus continuing to support forgetting as retrieval failure, available methodologies could not prove forgetting as storage failure.

1.1.2. Forgetting as loss of information over time.

A complementary way of describing forgetting, first used by Ebbinghaus (1885/1964), is as the assessment of retention of specific experiences over time. Slamecka and McElree (1983) emphasised that the fundamental distinction between retention and forgetting is that a retention measure is obtained from a single memory test, whereas forgetting is a measure derived from two or more retention tests separated in time. From a methodological point of view, the typical way of assessing forgetting is either to have separate groups exposed to the same information and test each of them at various delays; or to have the same group exposed to different types of information and counterbalance the type of information which is tested at each delay across subjects (Roediger, Weinstein & Agarwal, 2010). Retention is then plotted across the various time points and a forgetting curve is derived, which typically shows a loss of information as a function of time since encoding. With these designs, material is tested only once, as testing the same material repeatedly may alter the forgetting curve. Several studies have shown that retested material shows less forgetting than non-retested material (e.g., Roediger & Karpicke, 2006), suggesting that retesting reduces forgetting for the retested material. In contrast, as Ebbinghaus (1885/1964) put it: ‘Left to itself every mental content gradually loses its capacity for being revived, or at least suffers loss in this regard under the influence of time’ (p. 4).

When defining forgetting, there are some critical points that should be considered by any theories of memory (Cubelli, 2010):

A. To date no theoretical explanations are able to distinguish the underlying mechanisms of forgetting in neurologically healthy individuals and of patients with amnesia. Typically, amnesia is viewed as a more severe form of forgetting, differing only quantitatively, this view thus far lacks evidential support.

B. Though the existence of different memory components of memory is well known, forgetting has typically been investigated as a unitary phenomenon, rather than specific to each system (Wixted, 2007). Yet different forgetting curves have been obtained when using different memory tasks. For example, learning tasks produce rapid forgetting initially after learning, followed by later slower rates (e.g., Ebbinghaus, 1885). Conversely, for autobiographical memory, the majority of the recalled memories belong to the recent past, however, there is a rise in the curve for recalled memories of events near the age of 20 (e.g., Rubin, Rahal, & Poon, 1974). Lastly, short-term memory, when assessed with the Brown–Peterson test, produces rapid forgetting with nearly all information being lost after a short interval (Keppel & Underwood, 1962). Thus, forgetting would appear to be different in different memory systems. Consequently, different underlying mechanisms should be assumed to explain why information is no longer available when memory is tested.

C. Forgetting is typically viewed in a negative way, revealing memory failures, while remembering always is desired: the more information remembered, the better functioning is the memory. Case reports of exceptional ability to remember all information encountered, even irrelevant and unwanted details have, however, been associated with profound difficulties in everyday life. This suggests that forgetting serves a very positive role in avoiding memory becoming overloaded with the retention of large amounts of irrelevant detail, even to the extent of being described as a 'human superpower' (for a detailed discussion see Logie, Wolters, & Niven, 2018).

While the 3rd point made by Cubelli (and by Logie, Wolters, & Niven, 2018) is beyond the scope of the current thesis, the following sections of this chapter will address the first two points in more detail.

1.1.3. Forgetting Theories.

Two explanations for non-pathological forgetting have been proposed: the first assumes decay of memory traces, the second assumes interference between aspects of the memory trace (Hardt, Nader, & Nadel, 2013). The question of which of the two theories better explain forgetting commanded a great deal of attention early in the twentieth century (e.g. Mueller & Pilzecker, 1900; Jenkins & Dallenbach, 1924; McGeoch, 1932). Current consensus typically favours the second of the two (Roediger et al., 2010), proposing that interference is responsible for much of everyday forgetting, while the decay theory has generally been rejected for long-term forgetting (Roediger et al., 2010; Neath, & Brown, 2012). Similarly, with regard to short-term forgetting, the two main theories assume either: 1. temporal forgetting, through decay over time (e.g., Barrouillet, Bernardin, & Camos, 2004; Burgess & Hitch, 2006) or as a result of traces becoming less temporally distinct, thus becoming more difficult to retrieve with the passage of time (e.g., Brown, Neath, & Chater, 2007) or 2. theories suggest that forgetting results from subsequent events that cause interference (e.g., Nairne, 1990; Farrell & Lewandowsky, 2002). While interference has been favoured as the primary cause of forgetting in short-term memory, there have been several memory researchers arguing at least some involvement of decay processes (Radvansky, 2015). With recent investigations suggesting that forgetting in short-term memory is likely to occur due to a combination of decay and interference from other material (e.g., Altman & Schunn, 2012). A thorough discussion on the causes of forgetting in short-term memory is beyond the purpose of the current thesis (for a review see: e.g. Reitman, 1971; Ricker, Vergauwe & Cowan, 2016).

1.1.3.1. Decay theory.

The decay theory assumes that forgetting occurs as a result of passive, time-based, disruption of memory traces (Lemaire & Portrat, 2018). Thus, forgetting is a consequence of disuse, over time memories are weakened to the point where they become unrecoverable. Later research, however, proposes that while forgetting does occur with the passing of time, time itself is not the only cause of forgetting (Storm, 2018). A strong case against decay theory was put forth by McGeoch (1932) arguing its unsuitability as a scientific theory, as no mechanisms were provided to account for it. He additionally pointed to the experimental evidence coming from reminiscence studies (e.g., Brown, 1923) whereby a memory may not be accessible at one point but may be recovered at a later point, making it inconsistent with decay theory. Lastly, Jenkins and Dallenbach's (1924) experiments provided evidence for interference theory by showing that less forgetting of information occurred after sleep periods than after equivalent wakefulness. McGeoch (1932) proposed that even when passage of time was controlled for, forgetting could be determined by the number of events during that time, with more events causing greater forgetting.

Recent research proposes a further hypothesis, that the characteristics of forgetting depend on the underlying declarative memory representations (Sadeh, Ozubko, Winocur & Moscovitch, 2014). It is assumed that some memories depend on recollection, and these are more vulnerable to decay, while other memories depend on familiarity, and these are more vulnerable to interference. Recollection involves conscious awareness of an event and its contextual information and is said to be supported by the hippocampus. Hippocampus dependent memories, while relatively resistant to interference, are sensitive to decay. Familiarity on the other hand involves a feeling of "deja vu", void of any contextual information and is said to depend more on extrahippocampal structures, which are not resistant to interference. Therefore, forgetting will occur due to decay or interference depending on the nature of the encoded memory that influences whether retrieval is based on recollection or familiarity. This hypothesis is further explored in Chapter 7.

1.1.3.2. Interference theory.

One of the first demonstrations of forgetting as a result of interference was provided by Mueller and Pilzecker (1900). They found that if retrieval cue items used at test somehow became associated with another memory during the retention phase, then participants were less likely to recall the initial item on a later test. Thus, Mueller and Pilzecker's results supported interference theory, providing evidence that not time, but rather, the changes which occur during the passage of time, such as the encoding of new memories would cause forgetting. For example, today we can remember what we did yesterday evening, but after a few months this memory is likely to be forgotten, not because so much time has elapsed, rather due to the cluttering of memory for subsequent similar events. The cluttering makes the retrieval of any particular memory hard. Forgetting would occur not because the memories decay, rather due to the ever-changing structure of memory and due to its limitations in differentiating similar traces (Anderson, 2003).

The view that interference is a powerful cause of forgetting has stood the test of time. After many years of research, and many papers on the topic, interference based forgetting still posits that new information (presented prior or after the learning event) attenuates memory expression. Neuropsychological research has shown that consolidation processes start during learning, which then lead to the formation of stable memories in the long-term (McGaugh 2000; Squire, Genzel, Wixted, & Morris, 2015; Bailey et al., 2016). Before this stabilisation takes place, the new traces are labile and thus susceptible to disruption, either from new learning or interference (Wixted, 2004). Animal models allow for an investigation into the disturbance that is found on retention when electroconvulsive shock is given after training, which are well explained by consolidation theory and interference hypotheses (Adams & Peacock, 1965). Studies have found that an amnesic effect arises as a result of electroconvulsive shocks being administered subsequently after learning of a particular task. Specifically, this reduces the performance of animals on that task in subsequent trials. Several hypotheses have been proposed to account for this phenomenon. One hypothesis assumes that shocks disrupt the neural activity necessary for the consolidation of memory traces (Duncan, 1949). Another suggests that because of the aversive nature of electroconvulsive shocks the animals learn to avoid activities that are followed by them (Coons & Miller, 1960). Lastly, electroconvulsive shocks may elicit competing responses that interfere with acquisition and/or retention of other responses (Adams & Peacock, 1965).

Forgetting resulting from interference can be caused either proactively or retroactively. For example, when two lists are learned sequentially, a list A followed by list B, this may result in two types of interference. Retroactive interference, where the learning of list B disrupts retention of list A, or proactive interference where previous learning of list A disrupts learning (or remembering) of list B. When investigating retroactive interference, it is particularly difficult to disentangle the effects of the strengthening word pairs from list B (e.g., Dog-Sky) from the suppression of list A responses (e.g., Dog-Rock). This difficulty lies in the fact that the word pairs from list B receive additional strengthening through repeated study/test cycles. The difficulty when investigating list strengthening effects lies in disentangling the strengthening of one half of the word pairs from the output interference that they cause on later test (free recall) for the remaining non-strengthened words. In free recall tests, if participants are left to recall the list words in any order, it is typical for them to begin with the strengthened words, this output ‘order’ is likely to produce inhibition for the remaining words. Similar difficulties arise in the part-set cuing literature, where providing part-set cues often create overt or covert output interference biases (see Roediger & Neely, 1982; Nickerson, 1984; Anderson & Neely, 1996, part-set cuing and related research is further discussed in section 1.2.1 and Chapter 2).

In proactive interference the retrieval of consolidated long-term memory items can be hindered at the retrieval stage (Skaggs, 1933). For example, at the retrieval stage, competing items may interfere during recall. Though it was previously believed that this type of proactive or output interference determined whether or not a memory was retrieved (Dewar, Cowan, & Della Sala, 2007), more recent insight into retrieval processes suggests that retrieval may also affect the memory content itself (Hardt, Einarsson, & Nader, 2010). The act of retrieving a consolidated memory produces plasticity of its respective traces, its re-consolidation can be affected by later exposure to new material, similar to processes which happen after initial encoding (Nader & Hardt, 2009). This can result either in the incorporation of new material into the original memory or in certain instances decrease retention (Walker et al., 2003; Hupbach et al., 2007).

The effect of the act of retrieval or testing on retention has received much interest of late. The literature discussing the advantages of testing on retention (e.g., Roediger & Karpicke, 2006; Rowland, 2014; Karpicke, 2017; Greving & Richter, 2018) show that recalling information promotes its long-term preservation. The opposing literature, however, notes that recall can at times also have a negative effect on memories. These issues are discussed in the following section.

1.2. The effects of testing on retention.

Much of early literature assumed that future memory is improved only through study, but not testing. The effects of testing on retention were difficult to uncover due to limitations in methodological design and analysis. Yet, recent research has begun to demonstrate remarkably strong effects of testing, showing that remembering affects subsequent learning and retrieval. In many instances testing produces beneficial effects on memory performance, but in some instances, testing can have a harmful effect. This chapter next reviews and discuss both.

1.2.1. Retrieval induced forgetting and remembering.

Research from the last two decades has shown that retrieval, or testing, of a specific memory (i.e., retrieving part of a previous holiday) typically attenuates retrieval of other memories (i.e., memories for other details of that event) causing retrieval-induced forgetting (RIF; Anderson, Bjork, & Bjork, 1994). More recently, however, it has been shown that retrieval can both attenuate and aid recall of memories (Bäuml & Samenieh, 2010). Several studies have shown that testing can benefit long-term retrieval of both the material that is tested, and material that was initially presented but not tested (Chan, McDermott & Roediger, 2006; Carpenter, Pashler, & Vul, 2007; Pilotti, Chodorow, & Petrov, 2009; Chan, 2009; Thomas et al., 2018; Baddeley, Allen, Atkinson & Kemp, 2019). Other researchers suggest that retrieval of non-tested material is hindered by repeated retrieval of the tested material (Tandoh & Naka, 2007; García-Bajos, Migueles, & Anderson, 2009), and still others suggest that testing makes no difference to retrieval of the non-tested material (MacLeod & Macrae, 2001).

Retrieval induced forgetting (RIF) studies typically employ some form of retrieval-practice paradigm that consists of four phases: study phase, retrieval practice phase, distractor phase and final test phase (Chan, 2009). In this paradigm participants first study a list of categorised words (e.g., fruit: orange, cherry, banana; scotch, gin, rum, etc.), they then perform cued recall retrieval practice for half of the items from half of the categories (e.g., fruit - ch__; fruit - or__; but not any items in the drinks category). This is followed by a distractor phase, after which participants' memory performance is assessed with a final test. The paradigm denotes practiced items (e.g., cherry, orange) as Rp+, non-practiced items from the practiced category are denoted Rp (e.g., banana), and the items from the non-practiced category are denoted Nrp (e.g., scotch). The typical finding is Nrp items are recalled better than Rp items, suggesting that the retrieval of Rp+ items impairs recall of related Rp items. The mechanisms behind RIF, pertaining to what causes non-retrieved items to become less recallable, are not fully understood. Depending on the particular population and method under investigation, various factors have been proposed to influence forgetting, such as: response competition, strategy disruption, cue overload, cue biasing, or context biasing (for a review see Anderson & Bjork, 1994). A popular explanation is that retrieval strengthens the practiced items, which then interfere with, or hinder activation, of the non-practiced items that shared the same retrieval cue (Storm & Levi, 2012).

In contrast, recent research proposes that much of the previous literature investigating RIF may have missed an important aspect of retrieval. That of retrieval serving more than simply reinforcing memory of a tested fact, as retrieval can also improve recall of non-tested material (Chan, McDermott, & Roediger, 2006; Callender & McDaniel, 2007; Carpenter, Pashler, & Vul, 2007; Wirth & Bäuml, 2020). This phenomenon has been termed retrieval-induced facilitation (RIFA) (Chan, McDermott, & Roediger, 2006). This effect is described in a study by Carpenter, Pashler, and Vul (2007) who found that retrieval practice of target words enhanced subsequent recall of cue words when using a paired associates learning task. Participants in their experiment were exposed to either a retrieval practice or a restudy condition. After having learned word pairs (e.g. angle – corner), participants who took a cued recall retrieval test (e.g. angle –) had better performance on a delayed recall of the cue word (angle), relative to those who restudied the entire pair. Because the authors controlled for different exposure times to the cue word, by equating them between retrieval practice and restudy conditions, they conclude that the enhanced recall of the cue could only be attributed to retrieval practice of the target. Another example of RIFA is reported by Chan, McDermott and Roediger (2006). Subjects in their experiment were allocated to either a testing condition or a control condition. They either studied an article about toucan birds and then performed a cued recall test on that article or were dismissed after study. All participants took a final test

24 hours later. The test included both questions that appeared during retrieval practice (Rp+) and questions related to the ones that appeared during retrieval practice (Rp). For example, a Rp+ question/answer would be ‘Where do toucans sleep at night? /tree-holes’ and then a related Rp question/answer would be ‘What other species are related to toucans? /woodpeckers’. The study showed that the subjects in the testing condition outperformed subjects in the control condition, on the Rp items. More importantly this effect was not seen when subjects only restudied (without retrieval) the Rp+ items showing that retrieval is essential in producing this enhancement in performance.

Seemingly, the findings in the literature appear incompatible, some showing retrieval-induced forgetting and others retrieval-induced facilitation, though using similar manipulations (i.e., retrieval of a subset of initially studied material). Two of the critical factors that have been proposed to influence whether testing may have a positive or negative impact on later recall of the non-tested materials are delay (Tandoh & Naka, 2007; Chan, 2009; García-Bajos, Migueles, & Anderson, 2009; Baddeley, Allen, Atkinson & Kemp, 2019) and integration (Anderson et al., 2000; Bäuml & Hartinger, 2002; Chan, 2009; Baddeley, Allen, Atkinson & Kemp, 2019; but see e.g., Bäuml, 2019; Wirth & Bäuml, 2020 for alternative accounts).

1.2.2. Delay.

The negative effect of retrieval, RIF, is most usually found in studies that employ short time intervals between encoding and test, typically of 5 minutes or less (e.g., Anderson et al., 1994; Anderson & Spellman, 1995; Hicks & Starns, 2004; Jonker et al., 2013). This differs strikingly from people’s typical remembering in daily life, whereby remembering often takes place after longer time intervals and may occur at a different spatial location with respect to the encoding of the event (Pastötter & Bäuml, 2014). An early finding, that motivated future studies on the testing effects, was that repeated test trials produce equivalent amounts of learning as do repeated study trials (Tulving, 1967). Deriving from this finding, researchers set out to better understand the relationship between different learning methods (such as repeated testing) and delay. Although most studies of RIF are completed in the context of a single-session, some studies have examined the consequences of retrieval practice on final test performance using much longer delays, such as 24h or even a week (e.g., Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009; García-Bajos, Migueles, & Anderson, 2009: Saunders, Fernandes, & Kosnes, 2009; Storm, Bjork, & Bjork, 2012). For example, Roediger and Karpicke (2006) exposed participants to a final test either after 5 minutes or one week following initial study–test cycle and found significant interactions between learning method and delay. They found better performance for repeated study compared to repeated testing when the final test was taken after 5 minutes, however, performance was higher in the repeated testing condition if the final test was taken after one week (Karpicke & Roediger, 2006). Another landmark study that illustrates the effects of the interaction between delay and repeated testing was carried out by Spitzer (1939). He tested 3,605 students on 600-word articles at various delays across 63 days. In order to examine the effects of repeated testing on final performance, the study compared students who took a single test on the final delay (63 days later) to those who also took earlier tests. A forgetting curve was derived which showed that longer delays between encoding and final test produce lowered performance on that (final) test. The sooner the retest was carried out after encoding, the better the performance was on the later tests. For example, when comparing a group that was tested immediately after encoding and then a week later to a group which was not tested until day 21, when tested again on day 63, the first group showed much better performance. Additionally, because by day 21 forgetting had reached asymptote, the second group's performance was not enhanced at all. The conclusion of Spitzer's study was that the delay between encoding and first test must be relatively short in order to produce a positive effect at a later occasion.

Other studies that have examined the consequences of retrieval practice on final test performance using longer delays (weeks or months: e.g., Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; García-Bajos, Migueles, & Anderson, 2009; Saunders, Fernandes, & Kosnes, 2009; Storm, Bjork, & Bjork, 2012; Chan, 2009, 2012; Baddeley, Allen, Atkinson & Kemp, 2019) show that testing can benefit retention, not only of the tested but also of the non-tested (related) materials.

Theoretical implications for the effects of delay on RIF are still debated. Although some researchers have argued that the inhibition-based account predicts that RIF reflects a temporary or transient reduction in the accessibility of items in memory, expecting it to be diminished or even eliminated after a delay (MacLeod & Macrae, 2001; MacLeod & Hulbert, 2011; Raaijmakers & Jakab, 2013), others have argued that under certain conditions it might be possible for inhibition to have long lasting consequences (Storm et al., 2012; for further discussion see also: Anderson, 2003). Regardless, the temporal boundaries of RIF remain largely unknown, thus limiting our understanding of how retrieval practice affects the recall of other information in the long-term.

1.2.3. Material integration.

The suggestion that integration plays an important part in RIF comes from studies showing that competition between responses is eliminated when there are interrelationships between those responses (Anderson & McCulloch, 1999). RIF has been observed using different types of material (Tandoh & Naka, 2007; Little, Storm, & Bjork, 2011; Storm, Angello, & Bjork, 2011; García-Bajos & Migueles, 2013). An important factor that seems to determine whether RIF effects are or are not found is related to the nature of the material, specifically the degree to which participants can integrate the material to be remembered (e.g., Anderson & McCulloch, 1999; Bäuml & Hartinger, 2002). Material with high integration refers to complex knowledge structures composed of highly interconnected items (i.e., narratives, video clips), while low integrated material usually refers to lists of independent items (i.e., words, pictures).

For example, Tandoh and Naka (2007) showed that testing can impair delayed recall of the non-tested materials, using non-integrated materials. García-Bajos, Migueles, and Anderson (2009) used a video recording of an event and manipulated the retention interval between the initial and final test (immediate and 1-week delay). They found that highly integrated materials showed no RIF, but poorly integrated material did, even after one week. Chan, McDermott and Roediger (2006) found that partial testing of integrated material increased performance of the tested material, as well as the non-tested material. Chan (2009) investigated the relation between delay and RIF/facilitation, and showed, by using three retention intervals, that the influence of testing on the forgetting of the non-tested materials can be observed in both short and long-term memory. He concluded that testing can facilitate both tested and non-tested materials, and that this advantage is long-lasting (Chan, 2009). The facilitation effect appears to be sensitive to the level of integration and coherence embedded in the stimuli (Chan, 2009; Little, Storm, & Bjork, 2011). In the absence of such integration, testing can harm retention of non-tested but related information and RIF can occur (Anderson, 2003; Storm & Levy, 2012).

Given the contrasting results in previous studies, an understanding of the effects of integration requires further investigation. Moreover, since much of the research in retrieval practice stems from an interest in its implications for education (Leeming, 2002; Chan, McDermott & Roediger, 2006; Roediger & Karpicke, 2006; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel, Roediger, & McDermott, 2007), very few studies have investigated these issues in older and clinical populations.

Most of the evidence pointing to an enhancement in performance as a result of retesting, in older and clinical populations, comes from studies investigating practice and repetition effects in longitudinal clinical assessment. These studies have documented significantly higher benefits from retesting in healthy groups compared to those with suspected neurodegenerative disease such as Mild Cognitive Impairment (MCI) or dementia, notably Alzheimer's Disease (AD). Though the magnitude of this effect may be higher in cognitively intact individuals compared to those with MCI for some tests, it may be comparable for others (Heaton et al., 2001; Duff et al., 2008) and even may show greater benefit in MCI patients (Yan & Dick, 2006; Duff et al., 2008). While the literature with interest in education views retrieval induced facilitation effects as beneficial, the literature investigating this effect in clinical assessment have generally seen it as detrimental. Ferrer, Salthouse, Stewart, and Schwartz (2004) showed that when analysing longitudinal data across tests carried out repeatedly on multiple occasions, performance appeared not to change or even to improve across test sessions on some cognitive tests, a finding that was attributable to practice effects on the tests. For example, in individuals with MCI who have high risk of progression to AD, even a small increase in scores due to practice effects may obscure the underlying cognitive decline enough to delay the diagnosis by years (Duff, Horn, Foster, & Hoffman, 2015).

Baddeley, Rawlings and Hayes (2014) reported that young healthy people show little forgetting of verbal material (prose passages) when tested repeatedly, over delays of up to six weeks, while older participants exhibit faster forgetting. As the authors themselves pointed out, this result is not typical of extant literature. In a review of several studies, Salthouse (1991) only found evidence of faster forgetting for half of the older participants he tested. These findings have been confirmed in a meta-analysis by Kausler (1991). More recent individual studies also failed to find a difference in forgetting rates between younger and older participants (Ferrer, Salthouse, Stewart, & Schwartz, 2004; Andrés & Howard, 2011).

The understanding of how repetition will affect later performance and the underling mechanisms behind this effect are of great importance to theories of learning and memory, as well as for clinical diagnosis. Furthermore, the establishment of conditions such as accelerated long-term forgetting (ALF), as well as the need to design neuropsychological tests that are not affected by practice and are reliable for repeated assessment, amplify the necessity to understand the effects of repeated testing. Repeated testing by exposing participants to the entire material at each testing produces large practice effects, not only in healthy individuals, but also in clinical patients. It has been shown that this method of testing can mask evidence of ALF, as well as conversion from MCI to AD by obscuring subtle declines in cognition (Goldberg et al., 2015). Representing the entire material on each delay facilitates relearning and makes it difficult to disentangle learning from forgetting.

1.3. Accelerated long-term forgetting (ALF).

1.3.1. Defining ALF.

ALF describes the performance of an individual, or group, presenting with unimpaired learning of new information, but who subsequently forgets that information at an abnormally accelerated rate (compared with healthy individuals) over delays of days or weeks (Butler, Muhlert, & Zeman, 2010; Elliott, Isaac, & Muhlert, 2014). This phenomenon has also been referred to by other authors as long-term amnesia (Kapur et al., 1996; Kapur et al., 1997; Mayes et al., 2003), accelerated forgetting (Blake, Wroe, Breen, & McCarthy, 2000; Bell & Giovagnoli, 2007) or as ephemeral memories (Lucchelli & Spinnler, 1998). The use of the term ALF was intended to both distinguish the disorder from the amnesic syndrome as well as to include those cases where long-term memory performance is poorer yet not completely absent (Butler, Muhlert, & Zeman, 2010). Lastly, this term is used to acknowledge a clinical, and possibly also a mechanical, differentiation between forgetting of this type and the rapid early forgetting occurring in other neurological conditions (Butler, Muhlert, & Zeman, 2010). ALF is of great importance both theoretically and clinically. Clinically, ALF can go undetected by standard neuropsychological memory tests which were typically designed to test newly acquired memories after relatively short delays (about 30 minutes) as a result patients may remain undiagnosed and untreated. From a theoretical perspective, ALF could provide new insight into the processes underpinning successful long-term memory performance in healthy individuals.

1.3.2. ALF in clinical settings.

ALF (Blake, Wroe, Breen, & McCarthy, 2000), was first described by De Renzi and Lucchelli (1993) and has since been observed in a number of case and group studies, particularly in relation to individuals with epilepsy (see Butler & Zeman, 2008 and Elliott, Isaac, & Mulhert, 2014). Most studies of ALF involve adult patients with temporal lobe epilepsy (TLE) and transient epileptic amnesia (TEA), which is thought to represent a particular manifestation of TLE. It has been suggested that ALF may result in pathologies that involve seizures or sub-threshold seizure-like activity, resulting in disruption of long-term memory storage, or in individuals with underlying hippocampal pathology (Manes et al., 2008). Many studies have focused on and found associations between ALF and several seizure variables such as seizure frequency (Mameniskiene, Jatuzis, Kaubrys, & Budrys, 2006; Wilkinson et al., 2012), seizures that generalise (Narayanan, 2012), and subclinical discharges (Fitzgerald, Thayer, Mohamed, & Miller, 2013). Furthermore, ALF was resolved in some patients with TEA who received treatment with antiepileptic medication (Zeman, Boniface, & Hodges, 1998; Midorikawa & Kawamura, 2007).

Contrary to evidence for ALF being a seizure related phenomenon, Mayes and colleagues (2003) have proposed that ALF may arise when any of the structural components of the brain networks supporting long-term memory formation, or their interaction, are disrupted and that this disruption does not have to be caused by seizures for ALF to occur. Several lines of evidence support that ALF may not be seizure related. In a recent study by Lah and colleagues (2017) they show that a group of children with severe traumatic brain injury (TBI) presented with ALF for verbal material. Out of the 13 children assessed in their study, 8 presented with ALF. Although TBI participants were impaired, relative to the control group, on a standard delayed story recall (30 minutes test), the correlations between the standard delayed story recall scores and ALF were small and not statistically significant. Whereas the correlations between the recall test at the long (7 days) delay and ALF was statistically significant. A greater decline from the short 30 minutes test to the long 7 days test was associated with lower Glasgow Coma Scale score and diffuse subcortical injuries in their sample. The study supports the idea that ALF is not a seizure related phenomenon and raises the possibility that short-term and long-term memory systems may be independent (Lah & et al., 2017). Additionally, the study shows that standardised memory tests are not sensitive to ALF, a result which is of great importance for clinical work, suggesting that patients with ALF could potentially remain undiagnosed if assessed only on standardised tests.

A recent review of studies investigating long-term forgetting found evidence of ALF in neurological patients with no history of epilepsy, such as AD and MCI (Geurts, van der Werf, & Kessels, 2015). Of the 11 studies examining long-term forgetting in neurological patients only 3 found evidence of ALF. This discrepancy in results may be caused by differences in methodology between these studies (for a review see Elliott, Issac & Muhlert, 2014). Moreover, one study by Manes et al. (2008) has found evidence of ALF when assessing memory performance over a 6-week delay (but no impairments at standard delays) in a group of healthy older people with subjective memory complaints. The authors propose that, in some cases, ALF may be a precursor to neurodegenerative disease and could reflect mild disruption in memory systems outside the medial temporal lobes (e.g., retrosplenial cortex). Other investigations have provided results that some patients with MCI show ALF, which further adds to the possibility that ALF is a risk factor for conversion to dementia (Walsh et al., 2014).

Together, all these results indicate that there are some important clinical implications to consider. If ALF occurs in more than just a few cases in neurological disease, it may explain the mismatch between subjective and objective memory performance that is a frequent topic in the forgetting literature (Minett, Da Silva, Ortiz, & Bertolucci, 2008). This mismatch has often been attributed to anxiety or depression causing an alteration in poor self-perception of memory performance (Elixhauser et al., 1999; Piazzini, Canevini, Maggiori, & Canger, 2001). Yet, there is also evidence that subjective ratings of memory by healthy individuals do not correlate highly with objective measures of their memory performance (e.g., Hultsch, Hertzog, & Dixon, 1990; Arnold & Bayen, 2019). Nevertheless, ALF findings raise the possibility that these individuals are detecting a real memory impairment, which is not visible to standard neuropsychological tests. A recent review by Elliot, Isaac and Muhlert (2014) suggested that measuring forgetting over longer delays could be a more sensitive tool for assessing memory problems than over short delays. In their evaluation they underlined the need for the development and validation of testing materials and methodologies able to detect forgetting over the long-term. They discuss the methodological issues that can arise with this type of assessment and also highlight the theoretical importance of investigating accelerated forgetting.

1.3.3. Theoretical importance of ALF.

Studying patients with neurological disease has contributed to the current understanding of human memory. ALF may now provide new insight into the mechanisms underlying memory and long-term memory consolidation. From a theoretical point of view, ALF challenges the traditional concept of long-term memory formation involving a single stage system (Weingartner & Parker, 1984). If long-term memory formation did involve a single stage system then the disruption of short-term memory should automatically result in a deficit in long-term memory. ALF findings contradict this view, supporting theories which propose multiple short-term and long-term memory formation processes (e.g., Squire, Cohen, & Nadel, 1984; McGaugh, 2000; Kandel, 2001).

The pattern of memory in ALF, showing normal performance on standard tests probing recall or recognition at 30 minutes, together with impaired performance at long delays of days or weeks, suggests that ALF may be a disorder with impaired consolidation rather than impaired acquisition (Butler, Muhlert, & Zeman, 2010). The possibility of a subtle acquisition deficit cannot be entirely excluded, and requires further investigation (Butler, Muhlert, & Zeman, 2010). Results from ALF studies may provide further evidence for theoretical accounts suggesting that memory consolidation is an active and long-lasting process. Evidence supporting a long consolidation process also comes from findings on retrograde amnesia, where brain lesions produced much greater forgetting of recent memories for events occurring after the brain damage (anterograde) compared to memories for events that occurred prior to the brain damage (retrograde) (McGaugh, 2000). Therefore, the ALF results allow for the slow consolidation hypothesis to be investigated in an anterograde fashion, providing superior experimental control than previous methods. If ALF is caused by structural damage, this could imply that the damaged regions are essential in the consolidation process. On the other hand, if ALF is caused by seizures, it may suggest that the seizure activity in particular brain areas causes either the disruption of ongoing consolidation or erases already consolidated memories (Butler, Muhlert, & Zeman, 2010). While a precise time-frame is still under debate, it is likely that the consolidation process could take months or even years (Gold, 2006). Consolidation can be distinguished between two types: a fast consolidation process mediated by medial temporal-lobe structures; and a slow consolidation process mediated by the repeated activation of hippocampal-neocortical connections (Squire, 2009). The “standard consolidation theory” proposed by Squire and Alvarez (1995) assumes the independence of memories from the hippocampal region. Nadel and Moscovitch (1997) alternative “multiple trace theory” distinguishes between memory domains and assumes that episodic memories are dependent on hippocampal and neocortex regions while semantic memories are represented in neocortical structures (Nadel & Moscovitch, 1997; Sekeres, Winocur, & Moscovitch, 2018).

1.3.4. Methodological limitations in measuring ALF.

The assessment of long-term forgetting raises a number of methodological issues, starting with the need for matching patients and controls both in terms of demographic as well as cognitive variables, for those cases where patients also have impaired learning (by equating initial performance between them). A variety of techniques have been proposed for matching initial level of learning, each with its own limitations. For example, amount of exposure to the study material can be manipulated, however, studies have suggested that this may lead to “over-learning” and may mask early forgetting with a ceiling effect (Bell et al., 2005). Contrastingly, differing initial levels of performance may be used, followed by comparison of the overall shape of the forgetting curves. The problem when employing this method is that, thus far, no model of how variations in level of initial learning affect forgetting over time has been accepted (Rubin & Wenzel, 1996).

The nature of the tested material as well as of the retrieval task (free recall, cued recall, or recognition) is also likely to influence results. For example, an influential concept in the neuropsychology of TLE implies that verbal and non-verbal forms of memory are mainly segregated (localised to the left and right hippocampi, respectively). This view offers a frame for both pre-surgical decision making as well as the interpretation of postoperative outcomes (Saling, 2009) for operations to control epilepsy that result in hippocampal damage.

As discussed in earlier sections, forgetting may be different depending on the nature of the tested material, semantically related (e.g., a story) and unrelated (e.g., individual words), the frequency and way of testing (testing entire, or subparts) of material. The length of time between testing intervals needs to be chosen such as that it avoids both ceiling effects in control subjects and floor effects in patients. Most of these issues will be discussed in the following sections in more detail.

1.4. Forgetting in healthy populations.

1.4.1. Ageing and forgetting.

Changes in memory performance are some of the most common complaints of older adults (e.g., Hertzog & Dixon, 1994). While it is well known that memory performance declines with age (e.g., Ryan, 1992; Kausler, 1994), this decline in performance is not uniform across tasks (e.g., Burke & Light, 1981; Craik, 1983; Balota & Duchek, 1988; Shimamura, 1989; Schacter, Kihlstrom, Kaszniak & Valdiserri, 1993; Logie & Maylor, 2009; Maylor & Logie, 2010).

The presiding idea in memory research portrays human memory as a non-monolithic entity, but as one composed of different systems fulfilling different roles. Tulving (1983) identified five different systems: sensory memories, working memory, episodic memory, semantic memory and procedural memory. This distinction between memory systems is relevant to the discussion on forgetting across the lifespan, as differences in performance vary depending on the system under investigation. Significant age-related decreases in memory performance are seen in three main areas. First, impairments on working memory tasks (Zacks, Hasher, & Li, 2000) are frequently reported for older adults. These tasks typically require online processing to be carried out while simultaneously holding information in mind. Second, older people have difficulties in acquiring new information and are less prone to engage in elaborative encoding, these difficulties most likely arise due to inefficient processing related to lack of cognitive control. Finally, impairments in retrieving information from memory have also been frequently reported in relation to age, particularly in tasks that require effortful retrieval processing (i.e., uncued recall, PM tasks, source memory tasks). Certain aspects of memory performance, however, seem unimpaired in aging these include: motor learning, priming, some aspects of semantic memory, primary memory, some aspects of episodic memory (i.e., well known life events), recognition memory, and PM when measured in everyday life.

When discussing changes in memory performance, it is also important to distinguish between memories that involve intentional retrieval of previously stored information, and memories that manifest in subsequent behaviour without conscious awareness of these previous experiences. The former is declarative or explicit memory, which consists of memory for personally experienced events (episodic memory) and semantic memory consisting of “facts” about the world (Quillian, 1968; Rogers & McClelland, 2004). Most of the evidence pertaining to the differences in older adults’ performance on episodic and semantic memory tasks, suggest significantly larger disruptions in episodic memory (Balota, 1983). The latter, procedural or non-declarative memory reflects a broad number of phenomena, including memory of prior episodes or events, without explicit recollection of the past. Evidence towards the distinction between declarative and non-declarative memory is strongly supported by findings from the clinical literature (i.e., amnesiacs). Amnesiacs consistently show poor performance on declarative memory tasks (i.e., free recall, recognition), however, their performance on non-declarative memory tasks is relatively good (Squire, Bayley & Smith, 2009; Duff et al., 2020). Such dissociations between groups, regarding impairments in different memory processes have provided insight into the underpinnings of memory and understanding of normal memory functioning.

Lastly, further dissociations between age groups may become apparent with longer (and multiple) assessments between memory formation and retrieval. Do younger and older adults actually differ in the rate with which information is lost over time? Thus far, evidence on age-related differences in long-term forgetting is mixed (Elliott, Isaac, & Muhlert, 2014).

1.4.2. Faster forgetting for older compared to younger adults.

Forgetting is a frequent complaint by older people (Craik & Rose, 2012). Findings in age-related memory research relating to the rate at which information in long-term memory is forgotten have, however, been inconsistent. This point has been highlighted in two early reviews. Salthouse (1991) evaluated 22 studies, assessing older and younger adults' memory performance over several retention intervals, and found that exactly half showed similar rates of forgetting for younger and older adults, while the other half revealed higher forgetting rates for older adults (Wheeler, 2000). These findings were further confirmed in a review by Kausler (1991). He concluded that only around half of the studies assessed found faster forgetting in older adults. Wheeler (2000) argued that this pattern of findings in itself suggests that higher forgetting rates in older adults are highly probable but may be difficult to detect due to various methodological issues. A more recent review by Elliot, Isaac and Muhlert (2014) suggests that newer studies do generally find an effect of increased forgetting with older age. Six out of nine studies included in their review did find ALF with increasing age.

Several factors have been proposed to moderate age-related differences in forgetting rates: the delay between encoding and test (length of the retention interval), the type of test administered (i.e., recall or recognition), the type of material (visual, verbal) and the level of encoding achieved by participants. The effect of delay has been thoroughly reviewed by

Kausler (1994) and Salthouse (1991). They report conflicting results. With respect to short intervals the general pattern is that of equivalent forgetting for different age groups irrespective of the type of test (recognition, recall), while at longer intervals (24h or longer) older people typically show faster forgetting. Additionally, when recognition tests (but not recall tests) are used at longer delays, small or no age differences are reported (e.g., Rybarczyk, Hart, & Harkins, 1987). Recognition was deemed less sensitive to age compared to recall, in forgetting rates especially with longer delays as it is more resilient over time (Luh, 1922). It is generally acknowledged that older participants show impaired performance on tests of recall, that require conscious recollection of the learning experience, but not on tests of recognition that are usually supported by familiarity processes, which is relatively preserved in older adults (Tulving, 1985; Craik & McDowd, 1987).

Postman and Rau (1957) employed testing procedures comparing recognition and free recall and concluded that recognition is a process which is typically preserved in older adults (see also: Tulving, 1985; Perner & Ruffman, 1995; Wheeler, 2000). Even studies employing recognition tasks have provided mixed results. Harker and Riege (1985) found that older adults did not show accelerated forgetting on recognition tasks for both words and items over 20 minutes intervals despite having lower scores (especially for words) at 2 minutes testing. Mitchell, Brown, and Murphy (1990) found no age-related differences on a picture recognition tasks over longer delay intervals (3 weeks) but initial learning was not equated between older and younger groups. Rybarczyk, Hart, and Harkins (1987) also found similar forgetting rates for older and younger adults on a picture recognition task at 2 hours and 2 days after equating for initial learning. Park, Royal, and Morrell (1988) reported that older adults had similar performance to younger adults on a picture recognition test at 48 hours but showed accelerated forgetting over longer intervals (one week, 2 weeks). Huppert and Kopelman (1989) employed a picture recognition tasks and reported that older participants showed a mild acquisition deficit and higher forgetting rates at 1-day and 1-week intervals despite having similar performance at 10 minutes testing. Similarly, Park, Puglisi, and Smith (1986) reported faster forgetting by their older group at a 4-week delay (on a picture recognition task) despite equivalent performance on immediate test. An important issue related to their study is that participants were required to differentiate between original target pictures and distractor pictures but the same target and distractor pictures were used at both immediate and 4-week assessment (Elliott et al., 2014). This may have created source memory confusion in the older adults' group because remembering when information had been seen is significantly impaired in older people (McIntyre & Craik, 1987; Craik, Morris, Morris, & Loewen, 1990; Craik, Morris, Morris, & Loewen, 1990). Fjell and colleagues (2005) investigated age related differences in performance on visual recall and recognition tasks with tests after 30 minutes and 75 days. They found stronger correlations between performance and age on the final delay (after 75 days) compared to that on the first delay (30 minutes) for both tasks, but no significant age by delay interaction, concluding that their study provided no evidence in support of ALF in older adults.

Davis et al. (2003) compared performance on verbal recall and recognition tasks between 4 age groups (30–45 yo, 46–60 yo, 61–75 yo, 76–90 yo) at 2 retention intervals: 20 minutes and 1-day. On the verbal recognition task all 3 older groups showed accelerated forgetting compared to the youngest group at 1-day assessment, but only the eldest group showed accelerated forgetting at the 20 minutes assessment. When all participants where included in the analysis, on the verbal recall task the 3 older groups had lower performance compared to the youngest group at both retention intervals but there were no differences in the rate of forgetting. After selecting only participants matched at initial encoding, the eldest group showed accelerated forgetting at 1-day assessment. The authors concluded that their results are in line with findings by Kausler (1994) and Salthouse (1991) that age related differences in the rate of forgetting became apparent at longer but not at short intervals. They also proposed that, based on their results, though older participants did show an acquisition deficit, the faster forgetting rate is not just a reflection of poorer encoding but is rather 'qualitatively similar to the forgetting demonstrated by amnesic patients with hippocampal damage' (p. 1088).

Whilst the mixed findings in the literature on ageing and forgetting can be partially explained by methodological differences, another constant debate is whether, and to what extent, learning abilities and differences at encoding impact forgetting. Loftus (1985) claimed that differences at encoding represent an important methodological confound as they lead to scaling problems. Underwood (1954; 1964) found that once initial learning is equated, age-related differences in forgetting are not significant (Underwood, 1964). Yet, other studies do find accelerated forgetting in older groups even when younger and older groups' performance was matched at encoding (Wimer, 1960; Hulicka & Rust, 1964; Harwood & Naylor, 1969; Park, Puglisi, & Smith, 1986; Huppert & Kopelman, 1989).

Another factor that further complicates comparisons of forgetting rates between younger and older populations is the effect of testing itself. In the study of forgetting, multiple testing of the same individual may result in confounding effects that go beyond the initial encoding phase and can derive from retesting itself. Several studies have shown that older adults perform worse than younger adults with longer delays between encoding and test, but these age differences can be drastically diminished when an intervening test between study and final test is performed (Weeler, 2000; Fraundorf et al., 2019).

1.5. Forgetting in clinical populations.

1.5.1. Forgetting in dementia.

With increasing age, cognitive decline becomes associated with a number of neurodegenerative conditions, specifically the decline in memory performance. Some examples of such neurodegenerative conditions include AD, Lewy body disease, Parkinson's disease, cerebrovascular disease, and other neurological conditions that have been associated with memory impairments. Patterns of memory impairment depend on aetiology. Most commonly, brain injury affects memory for new information, yet it can also affect previously stored memories, such as in cases of severe TBI (Piolino et al., 2007, Esopenko & Levine, 2017) or semantic dementia (Nadel & Moscovitch, 1997; Hornberger & Piguet, 2012).

In AD or MCI, the pattern of memory impairments is characterised by the inability to encode new memories (Pike & Savage, 2008). Memory deficits are also common in stroke patients, however, their presentation is heterogeneous and depends on the location and amplitude of the stroke (Geurts, van der Werf, & Kessels, 2015). These patients may suffer from deficits in encoding and, or consolidation of different domains (Lim & Alexander, 2009; Saczynski et al., 2009). In the case of TBI patients, numerous memory aspects can become impaired, particularly effortful encoding and retrieval processes (Geurts, van der Werf, & Kessels, 2015), however, these patients seem to maintain a normal range of memory performance post TBI, contrasting with the case of severe amnesia patients (Vakil, 2005). Due to the frequency of memory problems, investigations of forgetting patterns are an essential part of neuropsychological assessment. Forgetting is typically assessed with neuropsychological tests that measure performance at delay intervals between 20 and 30 minutes between encoding and final test (Lezak, Howieson, Loring, & Fischer, 2004).

In many cases, encoding and storage impairments can be captured successfully using these standard neuropsychological tests (e.g., Carlesimo et al., 1995). For example, poor memory performance of AD patients is apparent even after short delays, information is lost at a steep rate in the first few minutes up to 1-hour following acquisition (e.g., Larrabee et al., 1993; Christensen et al., 1998; Reed, Paller, & Mungas, 1998; Kramer et al., 2004). Yet it is not uncommon for patients, whose performance is normal on neuropsychological tests, to still complain about memory issues. In contrast to AD, otherwise seemingly healthy individuals, who perform normally at immediate and at 30 minutes tests, exhibit accelerated forgetting after several days or weeks (Manes et al., 2008). It has been suggested that these patients exhibit similar long-term forgetting rates as patients with amnestic MCI (aMCI) indicating the existence of earlier stages in the processes evolving toward pathological aging (e.g., Mary, Schreiner & Peigneux, 2013). In consequence, in such cases, tests of long-term memory may be more sensitive in detecting memory problems (Butler & Zeman, 2008).

ALF was initially described in patients with epilepsy (Blake, Wroe, Breen, & McCarthy, 2000; Butler & Zeman, 2008). Studies on epilepsy added additional intervals, to those used in standard neuropsychological tests, varying between 24h and 6 weeks and as a result found evidence of accelerated forgetting. Lah and colleagues (2017) propose that structural lesions alone, to any of the brain network components involved in long-term memory formation, free of seizures, may result in ALF.

A recent review by Geurts, van der Werf, and Kessels (2015) found evidence of accelerated forgetting with long-term memory paradigms in AD and MCI patients. Other studies report a memory profile for aMCI similar to the profile described in the context of TLE and TEA. For example, Manes et al. (2008) investigated whether ALF could be an early sign of MCI. They compared forgetting rates between MCI patients, patients with subjective memory complaints and healthy age-matched controls. The patients with subjective memory complaints and the healthy controls had similar performances and outperformed the MCI patients on the immediate and 30-minute delay (recall of two short stories and the Rey complex figure). When performance was reassessed at the 6-week delay, both the MCI patients, and patients with subjective memory complaints had significantly worse memory performance than healthy controls, with the two patient groups becoming indistinguishable. Such similarities between profiles (e.g., MCI and TLE) highlight the importance to assess ALF in order to pinpoint early consolidation deficits, additionally these consolidation deficits may also be uncovered in normal aging populations. Though patients with aMCI do not meet the criteria for AD, they present with abnormal memory function for their age and education level, and have a high rate of conversion to AD. Therefore, accurate clarification of MCI cases will be highly dependent on the sensitivity and specificity of the tests used for evaluation.

1.5.2. Forgetting in AD.

AD is a degenerative brain disease and the most frequent form of dementia, comprising approximately 50-60% of cases (Barker et al., 2002; Wilson et al., 2012). AD is hallmarked by cognitive deficits that affect the person's ability to perform activities of daily living (Alzheimer's Association, 2016). While AD cognitive impairments vary across individuals, the earliest and most common initial symptom is memory impairment, specifically deficits in episodic memory (Rubin, Morris, Grant, & Vendegna, 1989; Peterson et al., 1994). The early involvement of the entorhinal cortex and CA1 of the hippocampus in AD supports the observed decline in episodic memory even at the initial stages of the disease (e.g., Braak et al., 1996). Yet, it has been proposed that measurable cognitive decline becomes apparent only after a substantial delay following pathological change (Weston et al., 2018).

There is clear agreement that AD patients perform significantly more poorly than healthy old people on tests of delayed recall, even minutes after presentation of the memory material (Welsh et al., 1991). Studies investigating whether AD patients present normal or accelerated forgetting have reported conflicting results. It has been suggested that some of these differences derive from methodological confounds.

An early study by Inglis (1959) exemplifies the problems that can arise in this line of research. The study investigated the contribution of impairments in learning and forgetting in a group of patients with dementia. He used a paired-associate learning test comprised of word pairs with varying degrees of difficulty. He concluded that the patients had severe learning deficits but also that they had retention problems. Yet, as part of his methodology there was a maximum of 60 learning trials. Healthy participants required between 10 and 14 learning trials to learn the new pairs, while the patients needed a mean of 54.8 to 59.2 learning trials. For the 30 minutes delay, the healthy participants required a mean of 3.9 trials to relearn the pairs, while the patients again needed a mean of 56.8 trials. Though, patients needed far more exposures to the material compared to the healthy controls, it cannot be excluded that dementing patients still did not learn to the same degree, as a substantial proportion of the patients probably had reached the maximum of 60 learning trials without achieving criterion performance for learning (Kopelman, 1992).

In an attempt to override such methodological problems Huppert and Piercy (1978) devised an experimental paradigm which allowed them to compare the temporal gradients of memory decay in various participant groups after equating for baseline performance. The task consisted of a yes-no picture recognition task which used photographic slides of pictures taken from magazines. They exposed amnesic patients four to eight times more to the learning material compared to controls, thus managing to match initial performance (at 10 minutes delay) between groups. They showed, in two consecutive studies (Huppert & Piercy, 1978, 1979), that forgetting rates of Korsakoff patients (on delay intervals of 24 hours and 7 days) were similar to that of controls. This finding, with the employment of the Huppert and Piercy's paradigm, was subsequently replicated in other studies (Squire, 1981; Kopelman, 1985; and Martone, Butters, & Trauner, 1986) and has also demonstrated normal rates of forgetting in AD patients (Corkin, Growdon, Nissen et al., 1984; Kopelman, 1985) with some potential exceptions (e.g., Carlesimo et al., 1995).

Kopelman (1985) using the Huppert-Piercy method investigated long-term forgetting in Korsakoff as well as AD patients over a 1-week delay. Initial performance was equated by using multiple exposures for the AD and Korsakoff patients (approximately 14 times more than that of the healthy controls). At a 10 minutes delay performance was matched between AD and Korsakoff patients, and their performance was only slightly below that of the control group. His study suggested that neither AD, nor Korsakoff groups, present with accelerated forgetting, as long as baseline performance is equated. Kopelman (1985) therefore concluded that the differences in performance between healthy controls demented patients, mainly reflect an acquisition impairment in the demented groups.

Carlesimo et al. (1995) used a modified version of the Huppert-Piercy method assessing yes/no recognition for pictures at 90 seconds, 10 minutes, 1 hour and 24 hours delay intervals. Participants in their study had to learn to a criterion of 80% correct. Those who failed to reach criterion after a first trial were re-exposed to the material only once and excluded thereafter. Thirteen AD patients, 8 vascular dementia patients, 9 amnesic patients (amnesia was developed as a consequence of TBI or stroke), 11 controls for the dementia groups, and 12 controls for the amnesic group were tested. Their results suggest that AD patient and amnesiacs have accelerated rates of forgetting compared to the controls (between 1h and 24h assessments) but that vascular dementia patients have normal forgetting rates. Yet, these findings should be interpreted with caution, as no statistical data concerning these conclusions was reported (Geurts, van der Werf, & Kessels, 2015).

More recent findings provide further support for normal rates of forgetting in AD. Degenszajn et al., (2001) assessed 15 AD patients with mild or moderate dementia and 15 normal matched controls with the Buschke Selective Reminding Test. Recall was evaluated at 30 minutes and at 24 hours after exposing all groups to six trials of learning. Their results showed poorer performance across the six learning trials for the AD patients as well across both delayed recall tests. The authors concluded that AD patients do not have higher rates of forgetting (Degenszajn et al., 2001).

On the other hand, findings from studies investigating autosomal dominant (familial) AD, suggest that ALF is an early presymptomatic feature of the disease (Weston et al., 2018). Autosomal dominant (familial) AD, which arises because of an inherited gene mutation, shares a number of pathophysiological and clinical features with the more common form of sporadic AD (Bateman et al., 2011). An important advantage for the study of presymptomatic cognitive changes is that autosomal dominant AD patients have predictable ages at symptom onset (estimated based on family history). Long-term forgetting rates of 21 AD mutation carriers and 14 controls was assessed by comparing 7 day recall with initial learning and 30 minutes recall on three tasks (list, story, and figure recall). They compared forgetting scores between mutation carriers and non-carriers (adjusting for age, IQ, and test set) and additionally assessed for association between ALF and estimated years to symptom onset (for mutation carriers). Weston et al. (2018) results suggest that ALF pre-dates other amnestic deficits in autosomal dominant (familial) AD and could potentially underpin the subjective memory complaints of AD patients.

An important but difficult to control for factor that needs to be considered when studying forgetting in amnesic or dementing patients is baseline performance, ensuring that adequate learning has been accomplished before attempting to measure forgetting rates. With few exceptions, a potential problem with the literature that has addressed the issue of long-term forgetting in clinical samples (as well as in the ageing literature) may arise from a failure to match performance on immediate test across groups (Andrés & Howard, 2011). Geurts, van der Werf and Kessels (2015) conducted a literature review only including those studies that have both equated baseline performance between the patient and the control groups and an immediate measure and a delayed measure (over 24h). They were only able to identify four studies meeting these criteria, two investigating AD and two investigating MCI. This leaves open the possibility that in a large number of studies the differences in forgetting rates resulted in part from encoding differences rather than differences in the ability to retain information over time.

1.6. Degree of initial learning and rate of forgetting.

There is still significant debate on how to compare forgetting rates between groups performing at different levels. Two main hypotheses have been proposed. The first suggests that initial degrees of learning do not influence the rate of forgetting in the long-term (Slamecka, 1985; Slamecka & McElree, 1983) while the second proposes that we cannot accurately compare forgetting rates unless initial learning is equated (Loftus, 1985a, 1985b).

Much of the literature in the 1970s and 1980s investigated how manipulations of exposure time and task difficulty may affect initial encoding and subsequently, rates of forgetting. For example, Slamecka and McElree (1983) proposed that initial levels of learning of healthy individuals affect the intercept across a range of tests (i.e., free recall, cued recall), but do not also influence subsequent forgetting rate. They reached this assumption by manipulating initial degrees of learning by exposing participants to either low degrees of learning (one study trial) or high degrees of learning (three study trials). Performance was assessed with free and cued recall on three delays (immediate, 1 day and 5 days). While varying the degree of learning did affect initial performance it had little effect on subsequent forgetting rates. Slamecka and McElree (1983) thus concluded that equating initial performance between groups is not necessary, as the course of forgetting is independent of degree of initial learning.

Loftus (1985a, 1985b) argued that if baseline performance significantly differs between groups, scaling problems affect comparison of subsequent forgetting. Loftus used an analogy based on the decay of radioactive material to describe the comparison of forgetting rates between individuals who start at different encoding levels. He proposed to imagine two chunks of radioactive material with different mass (one smaller and one larger) having the same half-life. The larger chunk will have a more rapid loss of weight than the smaller. In a similar fashion, groups having higher levels of initial learning will have more to forget. Another problematic issue in scaling relates to the level of difficulty of items (Keppel & Wickens, 2004). He proposed two ways in which scales could differentiate between groups with differing abilities, each with its own limitations. The first was to expose high performers to difficult items, however, this resulted in clustering low performers at the bottom of the scale. The second was to expose low performers to easy items, yet in this case high performers were clustered at the top of the scale. When groups differ on the initial level of learning and subsequently lose 'n' number of items from memory, it is assumed that the loss has the same meaning at both the top and bottom ends of the scale (Loftus, 1985b). Loftus believed that the loss of items with varying difficulty (easy vs. difficult items) reflect different degrees of forgetting (e.g., six difficult items reflect less forgetting than a loss of six easy items). This scaling problem leads to an underestimation of the rate of forgetting in the low performing group. In order to address this confound, Loftus used a different method, which involved the comparison over time of the horizontal distance between forgetting curves. This method recorded the time it took two groups to forget *n* amount of information. The assumption was that, with time, the two groups' forgetting curves would overlap. He used this method to analyse some of his previous data and concluded that higher degrees of original learning do lead to slower forgetting rates. Thus, reinforcing the notion that groups must be equated on initial learning.

An important yet under-examined finding in the literature relates to intra-group variability in forgetting rates: comparisons across individuals rather than between. This intra-individual variability was examined by Freed, Corkin, and Cohen (1987) in response to Huppert and Piercy's (1979) claim that while Korsakoff patients showed normal rates of forgetting, the amnesic patient 'HM' showed accelerated forgetting over one week, using similar matching procedures (equated performance at 10m). Freed, Corkin, and Cohen (1987) went on to assess HM on Yes-No and forced-choice recognition memory and compared his performance to that of seven healthy controls on both tasks. In order to get HM's initial performance to the same level of that of controls he was given four trials, as opposed to one, on equivalent sets of stimuli. HM had better performance than any of the controls for two of the trials on Yes/No recognition memory, and worse performance on the other two trials. Lastly, on the forced-choice recognition task, HM had better performance than controls on one trial, similar performance on a second trial, and on the last two trials his performance was as bad as that of the worst controls. Another important finding relating to these results is that different patterns of forgetting were found depending on how performance was measured, either by recognition or recall. Differences in rates of forgetting may vary, not only depending on differences in encoding, but also depending on the task that is being used. As an example, in contrast to the findings from recognition memory (Kopelman, 1985; Kopelman & Stanhope, 1997), when patients with temporal lobe, diencephalic, and AD pathology were tested on recall memory tasks they forgot faster than healthy participants (Frisk & Milner, 1990b; Kopelman & Stanhope, 1997; Christensen et al., 1998; Isaac & Mayes, 1999a, 1999b). An in-depth review on the differences between recognition and recall performance is beyond the scope of the current review.

An additional example, examining forgetting as a function of individual differences, is given by Kyllonen and Tirre (1988) by assessing the rate of item acquisition (name–number pairs) and general cognitive ability of 685 young military recruits. The participants in their study learned 13 paired associates and were subsequently administered a retention test. They equated initial learning by dropping pairs from the study list after participants gave a correct response on each. They found differences in forgetting rates between participants who learned pairs faster and those who learned them slower. Specifically, fast learners forgot fewer item pairs over time compared to slow learners, though the former received fewer learning opportunities. The authors also assessed whether the association between retention and ease in learning the item-pairs was mediated by other cognitive factors. This later analysis showed that cognitive factors such as general learning speed, knowledge, working memory, and reasoning did not have a direct effect on forgetting rate, independent of the individual differences in the ease of learning item-pairs. They conclude that rate of acquisition may influence forgetting rate.

A more recent investigation into differences across people, and not the examination of means across experimental conditions, is that of Zerr and colleagues (2018; see also Nelson et al., 2016). They examined individual differences in forgetting rates by sampling 281 participants who had to study 45 Lithuanian–English word pairs (Zerr et al., 2018, Study 1). After having learned the pairs, participants took an immediate cued-recall test with corrective feedback. If items were not recalled they then took a second test (only on the non-recalled items). This testing procedure was repeated until each of the pairs had been recalled correctly (only once). All participants then restudied all 45 pairs one last time before the final cued-recall test, on all items. The authors had three dependent measures of interest initial performance, the number of tests it took for a participant to learn all items and performance on the final test. Consistent with Kyllonen and Tirre's (1988) pattern of results, the highest predictor for final test performance was the number of tests to criterion (an index of a person's speed of learning). To conclude, fast learners tend to outperform slow learners from the beginning, they are quicker to reach criterion (by definition, as this is how fast learners are typically defined), and most importantly they retain this information better as seen on final test performance (Zerr et al., 2018). Additionally, this retention advantage for fast learners is seen also when learners are tested after long delays of days (Nelson et al., 2016) and weeks (Zerr, 2017).

In summary, there is a long history in assessing the role of different degrees of learning on subsequent rate of forgetting. The literature has revealed that there can be considerable variability in performance both within groups (e.g., Kopelman, 1985) and within individuals (e.g., Freed, Corkin, & Cohen, 1987), the latter could prove more important, or even disguise, differences between groups. If initial performance is matched across groups, it is particularly important to avoid ceiling and floor effects. Additionally, the issue of matching encoding, by exposing a group/or an individual to multiple learning trials, is likely to provide opportunities for re-consolidation of material at retrieval (Kopelman & Stanhope, 1997; Isaac & Mayes, 1999a, 1999b; Jansari, Davis, McGibbon, Firminger, & Kapur, 2010). Lastly, there can be differences in findings when performance is assessed with recall or recognition (Kopelman & Stanhope, 1997; Christensen et al., 1998; Isaac & Mayes, 1999a, 1999b).

1.7. Forgetting in different memory systems.

Much of the literature concerned with the study of forgetting, particularly on cognitive aging, focused on retrospective memory (RM), or memory for past events (for a review, see Light, 1991). In many instances, however, subjective memory complaints relate to the forgetting of planned intentions. 'The ability to remember to carry out intended actions in the future' is termed prospective memory (PM, Brandimonte, Einstein, & McDaniel, 1996). PM is responsible for many parts of everyday cognition, and failures in PM can be as debilitating as those in RM. Several studies (e.g., Crovitz & Daniel, 1984; Kliegel & Martin; 2003) have reported that PM accounts for 50 to 80% of daily forgetting. PM is responsible for the accomplishment of delayed intentions (Ellis, 1996) and is distinct from working memory, which involves preserving small amounts of information over a brief period of time

(Baddeley, 1986) and from RM, which involves the retrieval of episodes experienced in the past (Tulving, 1983).

Forgetting in PM can occur in different ways: forgetting to perform an intention, forgetting the content of the intention, and/or forgetting if an intended action has already been performed or not. There is, however, a relationship between retrospective and PM tasks, and this relationship functions in both directions. Specifically, future behaviour also involves a retrospective component, and past behaviour is related to subsequent behaviour directly - via a non-conscious process usually termed habit (McDaniel & Einstein, 1992; Cohen, West, & Craik, 2001). The former implies that in order to successfully perform a PM task, one requires not only to recall something that is to be done in the future, but also to retrieve what it is that needs to be done, and this latter component clearly implicates RM. According to the latter, behaviours which have been performed many times in the past are encoded in memory such that environmental cues serve to automatically elicit the behaviour.

When distinguishing between PM and RM tasks three main characteristics need to be considered: a PM task is future oriented; a PM task is embedded in the ongoing activity, which subsequently needs to be interrupted for the PM task to be performed; the PM intention/response requires self-initiation. For a PM task to be successful one needs to self-initiate retrieval of an intended action at a specific moment, while a successful RM task requires externally prompted retrieval of past information (Ellis & Kvavilashvili, 2000; Kliegel, McDaniel, & Einstein, 2007). It is precisely this last requirement that makes PM memory task distinct from other types of memory. Therefore, PM is often referred to as the ability to 'remembering to remember' (Scullin, McDaniel, & Shelton, 2013). During the period when a future intention is formed and the moment when it has to be executed, it is common for people to be engaged in various other activities, typically unrelated to the intention. It is therefore unlikely that the intention could be maintained in conscious awareness across very long intervals (Dismukes, 2010). Therefore, two of the most important issues central to PM research were to uncover the cognitive processes that support self-initiated retrieval of an intention at the right time and what are the factors responsible for PM failure. These two issues are central, not only for theoretical purposes, but also to the development of strategies for improving PM performance.

Several researchers have argued that the term PM is misleading (Dismukes, 2010) as other cognitive processes besides memory, such as attention, planning, time management also support performance of a PM task. The processes recruited for successful PM performance depend on the type of PM task and may vary across the stages of a PM task. Under the broader term of PM, researchers have distinguished between several PM tasks, considering factors such as the type of cue that signals the necessity to perform a previously planned intention, the frequency of performing the PM task and the length of the interval between intention formation and PM task performance. Not only are different types of PM task supported by different cognitive mechanism but they also differ in terms of PM failures. Consequently, the approach to devising techniques to improve PM performance has to consider the cognitive processes and types of failures specific for each type of PM tasks. These aspects are discussed and investigated in Chapter 6 and 7.

Irrespective of their type, PM tasks involve the same three stages of memory processes as RM tasks do, namely encoding, storage and retrieval. Each of these stages contribute to a successful PM performance, although their relative impact depends on the type of PM tasks. For example, failure to perform the intended action in response to a target event may occur because a strong action-target event association was not formed during encoding. In such case, strategies for enhancing PM performance should focus on the encoding stage. On the other hand, when a PM action is performed repeatedly and becomes automatic, failures in PM may occur because we cannot remember whether we have already performed the action and thus erroneously repeat it. In such case, strategies for enhancing PM performance should focus on increasing memory for the performed action, during the retrieval stage. Different aspects related to these 3 stages of a PM are detailed in Chapter 6 and 7.

The observed age-related declines in various cognitive domains, such as memory, are assumed to result from reductions in efficiency of cognitive resources (Henry, MacLeod, Phillips, & Crawford, 2004). In the context of PM particular aspects of age-related decline in performance are still under debate: especially the contradicting view regarding whether or not PM performance is still impaired in older adults; when it is supported by spontaneous retrieval; and the discrepancy of findings in naturalistic vs. laboratory experiments. The PM literature, however, concedes that age-related differences in PM performance are moderated by the type of PM tasks. A more in-depth analysis of age-related differences in PM performance, with a focus on the different types of PM and their underlying processes is provided in Chapter 6 and 7.

Previous sections have discussed both the positive impact repeated retrieval has on future performance, and instances in which it affects it in a negative way (in RM). Similarly, repeated retrieval of an encoded event in a PM context can affect subsequent performance on such tasks in a positive, or negative manner. In PM repeated retrieval may occur in several forms: repeated performance of a previously intended action; repeated retrieval (rehearsal) of the association between the intended action and the context or the PM cue as an encoding strategy; repeated mental retrieval of the PM intention (either spontaneous or conscious) during the retention interval.

For example, common everyday life repeated PM tasks include: taking medication at the appropriate time, locking the door when we leave the house, turning off the gas after cooking. Successfully carrying out these activities is essential for independent and autonomous living (Woods et al., 2008; Schmitter-Edgecombe, Woo, & Greeley, 2009). A crucial distinction needs to be made when discussing repeated retrieval in the context of PM. Habits are generally formed by the repeated pairing of a behaviour with its context of performance (Sheeran & Orbell, 1999). Ronis, Yates, and Kirscht (1989) defined habit as ‘an action that has been done many times and has become automatic’. That is, it is done without conscious thought, as in the case of locking the door. It is precisely this aspect that could represent a potential source of failures in PM, leading to repetition errors (REs). REs in the context of PM are detailed in the following sections.

On the other hand, repeated retrieval can be beneficial to PM performance when it involves cognitive rehearsal, during encoding of an intention, as well as when the intention is refreshed tough rehearsal or priming during the retention interval.

1.7.1. Types of PM tasks.

1.7.1.1. Event based vs time based PM.

Retrieval cues are essential for PM performance as they signal the fact that some action has to be performed. A PM cue may be an event, a location, a specific time or time frame, or a person. Harris (1984) outlined the clear distinction between PM tasks that require to keep an appointment and those that requiring ‘to do one thing before or after another’. Drawing from this distinction between two categories of PM cues, Einstein and McDaniel (1990) have further classified PM tasks as event based (e.g., remembering to relay a message to a friend when in class) and time based (e.g., remembering to take medication at 3 p.m.). What makes these two types of PM memory task fundamentally different is how the PM cue is generated (Einstein & McDaniel, 1990). In event-based PM the cue signalling the PM response is generated externally, provided by the environment. In time-based PM tasks one needs to remember to perform a certain task at a given moment in the future in the absence of any cue, thus without any external support. Therefore, a time-based PM requires a higher amount of self-initiated processing in order to perform the PM response and more monitoring efforts (to keep track of the time when a PM response has to be performed) compared to an event based PM task. It is assumed that these two types of tasks are supported by different processes (Einstein & McDaniel, 1990; Cona, Arcara, Tarantino, & Bisiacchi, 2012).

Rabbitt (1996) proposed that what further increases the difficulty of a time-based PM task is that monitoring for the appropriate time when to perform a planned action constitutes a secondary task in itself that interferes with the memory for the prospective action/intention. Several researchers have further fine-tuned the distinction between various PM tasks. Kvavilashvili and Ellis (1996) proposed that intentions that have to be executed upon encountering a certain location may be different compared to intentions to be executed while performing a certain activity. Ellis (1996) proposed a distinction between time-based PM tasks that have to be performed at a certain time point and those that may be performed within a broader time frame. Whatever the distinction, it is clear that the cue is essential for PM performance.

1.7.1.2. Habitual vs. episodic PM tasks.

With regard to the frequency a certain PM task is carried out, the literature has distinguished between episodic and habitual PM tasks. The former, are performed only once or infrequently, while the latter, are performed repeatedly in a routine-like manner (e.g., Meacham & Singer, 1977; Meacham & Leiman, 1982; Kvavilashvili & Ellis, 1996; Einstein et al., 1998).

When a PM task is performed repeatedly this changes the nature of the task from episodic to habitual (Meier et al., 2014). Another clear difference between episodic and habitual tasks relates to how intentions are formed. Each time we plan an episodic PM task, such as paying a bill or sending a letter, we form an explicit intention, whereas in habitual PM tasks, the intention to perform an action is often implicit. We do not explicitly form the intention to switch off the gas after we prepared coffee or to lock the door when we leave home (Meacham & Leiman, 1982 cited in, Hicks, Marsh, & Russell, 2000).

Meacham and Leiman (1975, 1982) proposed that habitual tasks are easier to initiate as they are embedded in a series of activities usually performed in the same sequence. Therefore, this rich and ordered contextual environment provides the cues necessary to initiate the habitual task, often automatically. With time, as the habitual task is performed more frequently the execution becomes more automatic. Episodic PM tasks are performed less frequently and rely more on remembering to perform the action. As they are not part of a routine, episodic tasks frequently require the interruption of normal flow of activities. Therefore, episodic PM tasks may require more effortful processing (McDaniel & Einstein, 2000).

Another significant difference between the various PM task types is that, while PM failures in episodic memory are mostly due to omissions (participants do not perform the intention), failures in habitual PM usually consist of both omission errors and commission errors (as participants forget that they have performed the task and erroneously repeat it). Failures in event-based PM usually occur due to interference with the ongoing tasks. Namely, we form an intention to perform a certain action in the future but become involved in other activities and forget to perform the intended action.

Three different views conceptualising REs in habitual PM have emerged from literature. The monitoring view proposes that once a PM intention was formed, individuals use preparatory monitoring processes in order to preserve the intention in conscious awareness and to scan the environment for the target cue signalling that the intention has to be retrieved (e.g., Smith, 2003). REs occur when these effortful process are not deactivated once the intention has been performed. A second view proposes that intentions are activated in consciousness at a higher level than other memories (Goschke & Kuhl, 1993), and may be preserved after the intention was completed. According to Walser, Fischer, and Goschke (2012) the possibility for a no longer relevant cue to still elicit retrieval of an intention depends on two concurrent factors, the residual level of activation of the intention and whether preparatory monitoring processes have been disengaged or not. A third view proposes that PM intentions can be retrieved automatically without involving monitoring process (e.g., Einstein et al., 2005; Marsh, Hicks,

& Cook, 2006; Knight et al., 2011). Especially when there is a strong association between the intention and the target cue, retrieval of the intention upon encountering of the target cue may become automatic, thus facilitate REs. A more detailed discussion about PM failures in habitual tasks is provided in Chapter 7.

1.7.1.3. Short term vs long-term PM tasks.

As proposed by Baddeley and Wilkins (1984), similar to RM tasks, there is a clear distinction between short and long-term intentions, based on the length of the interval between intention formation (encoding) and the specific moment when the action/intention has to be performed (retrieval). The length of the retention interval significantly impacts the cognitive processes employed in PM tasks performance. Short-term intentions may be preserved in conscious awareness and remain active in working memory during the retention interval (Baddeley & Wilkins, 1984). In contrast, in case of long-term intentions the retention interval is always filled with multiple ongoing activities/tasks, that are typically not related to the PM intention. These activities draw on working memory and attentional resources, thus the PM intention can hardly be maintained in conscious awareness. The PM cue signalling that a certain intention/action has to be performed is embedded in the context of these array of activities and therefore is more difficult to be distinguished than in the case of short-term PM tasks. Thus, the realisation of long-term intentions is likely to be supported by different processes in terms of degrees of conscious monitoring (Kvavilashvili & Ellis, 1996). For the realisation of long-term intentions, conscious awareness may be relevant only in the proximity of the moment when the intended action has to be performed (McDaniel & Einstein, 1993; Ellis, 1996) or may be even irrelevant in PM tasks where retrieval is spontaneous/somehow automatic such as routine PM tasks (Goschke & Kuhl, 1996) or task where performance of the PM tasks may be triggered by the PM cue or by the context (Neumann & Klotz, 1994).

The effect of the duration of the retention interval in PM is not as clear-cut as in RM. It has been long established that RM declines over increased retention intervals (Ebbinghaus, 1885/1964). In PM, there is no consensus in literature regarding the effects of delay (retention interval) on PM performance. Several studies found a decline in PM response at increased delays (Loftus, 1971; Meacham & Leiman, 1982; Brandimonte & Passolunghi, 1994), other studies reported that PM response increases at longer delays (Hicks, Marsh, & Russell, 2000; Martin, Brown, & Hicks, 2011) and others reported no influence of delay intervals (Wilkins, 1976; Einstein, Holland, McDaniel, & Guynn, 1992; Guynn, McDaniel, &

Einstein, 1998). Conte and McBride (2018) found that monitoring for the cue and PM performance decreased with longer delays for event-based task but not for time-based tasks. These studies, however, used laboratory paradigms and compared relatively short retention intervals.

Another line of research found that during the retention interval, stored intentions are refreshed in memory either intentionally, through rehearsal, or spontaneously, through priming (Meacham & Leiman, 1982). Several studies reported that participants who thought about the PM task during the retention interval more frequently were more accurate in providing the PM response (Meier, Zimmermann, & Perrig, 2006). Meier, Zimmermann, and Perrig (2006) reported that presenting participants associated primes before the PM cues led to spontaneous activation of the intention and enhanced PM performance. Kvavilashvili and Fisher (2007) reported that rehearsing the PM task during retention intervals facilitated spontaneous retrieval. Holbrook and Dismukes (2009) investigated various aspect of PM performance in everyday life by requiring participants to keep a diary over a week. They found that 43% of the intentions were recalled spontaneously during this retention interval and were positively correlated with PM performance. Similar results were reported in previous diary studies (Ellis & Nimmo-Smith, 1993; Kvavilashvili & Fisher, 2007). These retention-interval recollections might increase the level of activation of the intention and might strengthen the associations formed during encoding (Dismukes, 2010).

1.7.2. PM and ageing.

A reduction in the efficiency of cognitive resources is assumed to contribute to age-related decline observed in a number of cognitive domains, including memory, particularly its prospective aspect (Henry, MacLeod, Phillips, & Crawford, 2004). Age related decline in PM has been the subject of a long-lasting debate that has yet to be settled. Craik's (1986) theoretical account proposes that age related decline in PM should be as significant as the decline in RM. Einstein and McDaniel (1990) proposed that PM is spared with age, thus being an ‘exciting exception to typically found age-related decrements in memory’ (p. 724).

Several recent reviews of PM literature have tried to reconcile these two views. Henry, MacLeod, Phillips, and Crawford’s (2004) meta-analysis reached the conclusion that age-related declines in PM performance were less significant when the PM task engaged spontaneous/automatic processes, and more significant when engaging more elaborate processes. They also concluded that the age-related effects reported in the PM studies reviewed were smaller than the effects in the retrospective literature, thus contradicting Craik's (1986) theory. This conclusion was later contradicted by Uttl’s (2006) meta-analysis who claimed that the real magnitude of age effects in PM studies may have been obscured by the fact that many studies suffered from significant ceiling effects in younger groups. An even more recent meta-analysis by Uttl (2008) showed that there is clear support in the existing literature that PM performance is affected by age, thus contradicting Einstein and McDaniel (1990) claim that PM may be spared with age.

Recent experiments directly comparing age related differences between PM task types, event based and time-based tasks, also provided mixed results. Rendell and Thomson (1999) reported impaired PM performance in older adults compared to younger adults for both time-based and event-based tasks. Kvavilashvili and colleagues (2009) found significant age effects in the time-based task but not in event-based tasks. Schnitzspahn and Kliegel (2009) contrasted performance between an event-based task and a time-based task in two older groups (young-old: 60–75 years and old-old adults: 76–90 years). They found significant age deficits for both types of tasks. Even though a firm conclusion regarding age related deficits in event-based PM task cannot be derived from the existing literature, existing meta-analyses propose that age related declines in older adults may exist irrespective of the type of episodic PM task (Henry, MacLeod, Phillips, & Crawford, 2004; Uttl, 2008).

When reviewing patterns of age decline in PM, Uttl (2008) concluded that, as resulting from laboratory studies, performance on event-based PM remains stable until 60 years of age and afterwards declines significantly. Several other studies have found that adults over the age of 60 do not constitute a homogenous age group and PM continues to decline significantly with age (e.g., Rendell & Thomson 1999; Kliegel & Jäger, 2006; Kvavilashvili et al. 2009; Schnitzspahn & Kliegel, 2009). Kvavilashvili et al. (2009) compared young-young (18-30 years), young-old (61-70 years), and old-old (71-80 years) performance on activity-based, event-based and time-based PM tasks. They reported that in the activity-based task age related decrements in performance were only found for the old-old group. On the time-based task younger participants outperformed both older age groups who performed at the same level. No age effects were found in the event-based condition. Schnitzspahn and Kliegel (2009) found age related decrements in PM performance between young-old (60–75 years) and old-old adults (76–90 years) in both event and time-based PM tasks. Kliegel and Jäger

(2006) compared PM performance between four age groups, a young-young (22–31 years), a young-old (60–69 years), a middle-old (70–79 years) and an old-old (80–91 years) but only the young-old age group different significantly from the younger group. Maylor and Logie (2010) conducted a large scale investigation between PM and RM between age groups from 8 to 50 years of age. Their internet study findings showed that developmental trajectories between PM and RM are qualitatively different and that PM cues differently influence performance at younger and older ages. All these results point to the necessity to define more groups, with a narrower age range, in order to avoid obscuring or enhancing age related deficits in PM performance (e.g. Logie & Maylor, 2009; Maylor & Logie, 2010).

1.7.3. Naturalistic vs. laboratory studies.

Birt (2013) reached the conclusion that, in general, laboratory studies show robust age-related deficits in PM while studies using naturalistic PM tasks usually report better performance for older adults compared to younger adults. Empirical studies suggest that older adults tend to perform as well, or better than their younger counterparts in time-based PM tasks that are carried out in naturalistic rather than laboratory settings. Naturalistic experiments usually require participants to either: telephone back at a specific time after a certain delay interval (either 4 weeks, 3 weeks or 5 days - Devolder, Brigham, & Pressley, 1990; Schaffer & Poon, 1982; Moscovitch, 1982 and Maylor, 1990, respectively); mail a postcard (Patton & Meit, 1993); or use an electronic organiser to log a specific time (e.g., Rendell & Thomson, 1993, 1999).

The magnitude of age effects on PM tasks is likely to depend on whether external aids are used to cue the PM event or not. Older participants use external aids more frequently in naturalistic tasks by setting up external cues to act as reminders. The fact that older adults rely more on external aids is not surprising since they more commonly report everyday memory failures and are also more concerned about them (Cavanaugh, Grady, & Perlmutter, 1983; Henry et al, 2004). Additionally, older and younger adults' motivation for completing PM tasks outside the laboratory may also differ (Patton & Meit, 1993; Rendell & Craik, 2000). The importance of an intention is subjective and depends on individual objectives, motivations and expected consequences (Walter & Meier, 2014). Studies investigating how task importance may impact PM performance have either manipulated the level of importance through instructions, providing a reward or increasing task attractiveness. A different approach was to require participants to score their perception of the importance of the task. In their meta-analysis Walter and Meier (2014) concluded that in general, importance of a task is positively correlated with PM performance (Meacham & Singer,1977; Kvavilashvili, 1987; Einstein et al., 2005). While others have found no significant benefits (Kliegel et al., 2001, 2004; Loft & Yeo, 2007; Brandimonte, Ferrante, Bianco & Villani, 2010).

This discrepancy in finding between laboratory and naturalistic studies was further supported by a more recent meta-analysis (e.g., Henry, MacLeod, Phillips, & Crawford, 2004) and was termed the ‘age PM paradox’ (Rendell & Thomson, 1999; Rendell & Craik, 2000; Schnitzspahn, Zeintl, Jäger, & Kliegel, 2011). Uttl’s (2008) meta-analysis suggested that the conclusions that older adults perform better that younger adults when PM is assessed in naturalistic experiments and poorer when assessed in laboratory setting is not entirely justified. The main arguments were that comparisons between the two types of experimental studies did not distinguished between studies using event-cued PM tasks and those using time-cued tasks.

The same conclusion was reached in a recent study by Schnitzspahn, Kvavilashvili and Altgassen (2018) who investigated age related differences in PM performance in time and event-based PM task in both types of experimental conditions (laboratory and naturalistic) within the same sample. Younger adults performed better than older adults on event-based tasks in laboratory but not in naturalistic experimental condition. Older adults outperformed younger adults only in naturalistic time-based tasks. These propose that age benefits on PM task performance reported in naturalistic studies may have been biased by the predominant use of time-based PM tasks.

1.7.4. Strategies at encoding in enhancing PM performance.

Several encoding strategies, aimed at enhancing PM performance, have been explored in the literature. For example, strategies implementing future thinking instructions, which consist of instructing participants to imagine themselves performing an intention, during encoding. Several experimental studies did find significant benefits of future thinking instructions on PM performance. Altgassen, Kretschmer and Schnitzspahn (2017) found that enhancing encoding of a PM task by asking participants to imagine themselves performing the task improved PM performance in both younger and older adults. Similarly, a future thinking manipulation during encoding benefited PM performance in a student sample (Neroni, Gamboz, & Brandimonte, 2014). The mechanisms underlying the effects of future thinking instructions on PM performance are still under debate. One hypothesis proposes that forming a visual representation of the future act may produce more durable memory traces of the intention by strengthening encoding. The alternative hypothesis is that future thinking encoding may produce a stronger cue-context association and elicit automatic retrieval of the intention (Paraskevaides et al., 2010).

Altgassen, Kretschmer and Schnitzspahn (2017) compared PM performance under 3 different encoding strategies, future thinking, repeated-encoding and simple encoding in adolescents, younger adults and older adults. Their results showed that future thinking instructions benefitted all age groups compared to simple instructions. Overall, PM performance was the best under the repeated-encoding condition, however, adolescents benefited more from future thinking instructions while younger adults from repeated-encoding. Seen as repeated encoding enhanced PM performance beyond future thinking instruction they propose that the most important mechanism underlying the effect of future thinking in PM performance is by creating stronger memory traces of the PM intention.

PM performance has also been improved using implementation intentions encoding (IIE) strategies (Gollwitzer 1993; Gollwitzer & Sheeran, 2006). Implementation intentions employ a deep encoding “if-then” strategy, “If Situation X is encountered, then I will perform behaviour Y!” (Brandstätter, Langfelder, & Gollwitzer, 2001, p.946). It was proposed that deep encoding of the cue-action association will automatically prompt the action /PM response upon occurrence of the cue (Einstein & McDaniel, 1990, McDaniel et al., 2004) and that ‘action initiation becomes swift, efficient, and does not require conscious intent’ (Gollwitzer 1999, p. 495). Miller, Galanter, and Pribram (1960), proposed that intentions are activated in memory at a higher level than memories that are not associated with an intention, thus the former are easier to retrieve in the future. Several mechanisms have been proposed to explain how IIE facilitates self-initiated retrieval of a PM task. One such mechanisms requires increasing the level of activation of the intent, by increasing the salience of the cue or by strengthening the association between the cue and target action to be performed (Chasteen, Park, & Schwarz, 2001; McDaniel, Howard, & Butler, 2008; Cohen & Gollwitzer, 2008; Zimmermann & Meier, 2010, Wieber et al., 2011). While IIE have consistently demonstrated improvement in PM performance in episodic PM tasks, the strategy may have a negative effect on performance in habitual PM tasks. IIE strategies may exacerbate the difficulty of ‘turning off’ a completed intention thus increasing commission errors (Bugg, Scullin, & McDaniel, 2013).

Most studies assessing different strategies to stimulate successful PM task performance have been carried out in laboratory setting and usually employed short delays intervals between encoding and retrieval of a PM task. In laboratory settings, the prospective step (‘when’) of the PM process typically happens over a relatively short interval. While in real life many PM activities cannot be performed so quickly, rather one often needs to form an intention that can only be carried out after days or weeks. Over long delays the relative contribution of retrospective and prospective components on successful completion of a PM task may change (Nigro & Cicogna, 2000), with the retrospective component becoming more relevant. In everyday life we often meet with situations in which we know that we are supposed to perform a certain action but cannot remember what the action is. Association with a former intent does not necessarily involve association with the specific action.

1.8. Objectives of the current thesis.

As reviewed in the previous section, it has become apparent that although we know a great deal about learning and retrieval, this is not yet the case for long-term forgetting. Among the various themes that require further investigation, the current thesis has chosen to address a few. The progress of forgetting over time will require repeated testing, yet standard clinical memory tests are typically designed for single assessment after periods of 40 minutes to one hour, making them unsuitable for assessing ALF. There is strong evidence that repeated testing may enhance recall, however, we also have evidence from part-set cueing studies that probing one item may reduce the memorability of others within the set. A potential solution is to use integrated material testing different samples at each delay. This does not enhance, but does delay forgetting, but how? (1) Possibly via a relearning effect in which case the patient with learning problems should show faster forgetting (2) A priming effect. Priming can be preserved despite dense learning problems. If an effect is due to priming existing memory representations, then it should be preserved in groups that show learning deficits. This will be explored in both healthy older adults (Chapter 2) and amnesic patient populations (mild AD; Chapter 3). Previous studies have suggested that both aging and AD impairs recall but does not disrupt priming. Hence, these groups should take longer to learn, but show reduced forgetting rates when tested with different feature probes at each delay. A second theme which will be addressed is that of degrees of learning. As discussed in the general introduction (Section 1.4), a common assumption in the field is that when comparing two groups they should be matched on the level of initial learning. This will be initially addressed in a replication study of a classic paper in the degrees of learning literature (Slamecka & McElree, 1983; Chapter 4), which was conducted with the help of a master's student. Unlike the data presented in Chapter 4, which assess between group performance, achieving a certain encoding criterion during learning, for the study of ALF, will require repeated exposure for low performing individuals, thus opening up the discussion on intra-individual variability (Chapter 5). Finally, much of the forgetting literature focuses on RM, and in the case of both healthy and pathological ageing, it has almost invariably reported substantial deficits in this aspect of cognition. Recent interest has started to shift to investigating PM, specifically forgetting of future intentions. The following experiments have looked at the effect of repeated retrieval on long-term memory, in the last two chapters the issue of repeated retrieval will be assessed in the context of PM. Specifically, the effect of repeated retrieval of the PM intention during encoding, used as a strategy to enhance long-term PM performance will be investigated in a sample of young healthy adults (Chapters 6). Repeatedly performing an action impacts memory performance, as in the case of habitual PM, how errors on such tasks can be reduced in both healthy younger and older individuals will be investigated last (Chapters 7).

CHAPTER 2: Assessing the effect of repeated partial testing on long-term memory with different types of material.

The first two experiments described in this chapter were designed to address the issues of how repeated partial testing affects long-term forgetting. If partial testing will have a different effect on memory performance depending on the population which is being assessed this could prove useful in informing us of the underlying mechanism behind this effect. Specifically, Experiment 1 and 2 compared the magnitude of the partial testing effect between younger and older participants utilising findings that ageing impairs episodic learning, while priming is preserved. The partial testing advantage should be reduced in the older group if it depends on relearning, but not if it depends on priming.

Secondly, it is equally important to determine if the benefits which arises as a result of partial testing hold with a variety of material. Chapter 1 has discussed that part-list cuing studies have typically found a detrimental effect of partial testing. Other lines of research have proposed that the nature of material being tested may partly determine whether testing inhibits or enhances subsequent forgetting (Chan, 2009). Specifically, retrieval induced forgetting may be eliminated if participants are able to integrate the studied material into a coherent and interconnected episode (e.g., Anderson, Green, & McCulloch, 2000; Bäuml & Hartinger, 2002; Migueles & Garcia-Bajos, 2007). Thus, in Experiment 2 this issue was assessed using a design that disrupts the overall theme of prose passages by scrambling the constituent sentences within and between passages.

Finally, understanding how partial testing operates may prove useful in the assessment of accelerated long-term forgetting, specifically in the design of tests that would be capable of accurately measuring forgetting over the long-term. Research has shown that repeated complete recall can enhance learning, thus leading to large practice effects in both healthy and clinical populations. Recall of specific items can impair retention of other material. A possible compromise has been recently proposed by Baddeley, Allen, Atkinson and Kemp (2019) to use integrated material and then test different samples at each delay. Their method does not enhance, but does delay forgetting, the following two experiments were designed to further investigate how.

Experiment 1: Sampled testing delays forgetting of unsampled items. Is it an effect of re-learning or of priming?

2.1. Introduction.

One of the most striking features of recent research on human memory has been the discovery that retrieval of material to be remembered can serve as a more powerful means of learning than its re-presentation in further learning trials (e.g., Roediger & Karpicke, 2006; Karpicke, & Roediger, 2008; Roediger & Butler, 2011). In a review of the literature Roediger & Butler (2011) have shown that retrieval practice (as occurs during testing) increases long-term retention, often to a greater extent than restudying the same material. The facilitatory effect of partial testing is in stark contrast with the large body of literature showing a detrimental effect of selective retrieval on non-retrieved items (e.g., Blaxton & Neely, 1983; Brown, 1981; Roediger & Schmidt, 1980; Roediger, 1978; Roediger, 1974; Rundus, 1973; Tulving & Arbuckle, 1966). This finding, has been referred to as retrieval-induced forgetting (for recent reviews, see Storm et al., 2015 and Bäuml & Kliegl, 2017). Retrieval-induced forgetting typically arises when one of two experimental tasks are employed: the output-interference task and the retrieval-practice task (Wirth & Bäuml, 2020).

The output-interference task examines whether the serial position of a studied item in the testing sequence influences its chance of being recalled. The item's probability of being recalled is dependent on output position and declines in accuracy as a function of the items' serial position at test (e.g., Tulving & Arbuckle, 1963; Smith, 1971), suggesting that selective recall of early items has a negative effect on the recall of later items. The retrieval-practice task typically uses a design where participants are asked to study a list of items, they then repeatedly retrieve a subset of items from the study list, finally on a later test they are asked to recall all studied items. The typical finding here is that, in comparison to a control condition without retrieval practice, repeated retrieval of a subset of items enhanced recall of the retrieved items yet it inhibited recall of the nonretrieved items, reflecting the retrieval induced forgetting effect (e.g., Anderson et al., 1994; Anderson & Spellman, 1995).

With both types of tasks, retrieval induced forgetting effects arise using a wide range of materials and experimental settings (for recent reviews, see Storm et al., 2015 and Bäuml & Kliegl, 2017). Yet, in nearly all retrieval induced forgetting studies, partial retrieval practice occurred shortly after the study phase, with delay intervals between study and test of 5 minutes or less (e.g., Anderson et al., 1994; Anderson & Spellman, 1995; Hicks & Starns, 2004; Jonker et al., 2013).

2.1.2. Retrieval of the whole set of material.

Over the past century, research has repeatedly demonstrated that retrieval is a potent memory enhancer (Abbott, 1909; Spitzer, 1939; Chan & McDermott, 2007; Karpicke & Roediger, 2008). This enhancement is known by various terms: test-enhanced learning, the testing effect, and in this study, the retrieval practice effect (Rickard & Pan, 2018). Studies examining the retrieval practice effect on later memory performance have shown that when practice tests involve recall of the entire material, its beneficial effect on memory is generally greater than when restudying it. This finding has been replicated in many experiments, in both laboratory and classroom settings, using various types of materials (i.e., foreign language vocabulary, images, general knowledge questions, prose) and in various subject populations (e.g., Gates, 1917; Spitzer, 1939; Roediger & Karpicke, 2006; Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; Butler & Roediger, 2008).

For example, Roediger and Karpicke (2006) tested participants, in different conditions, on free recall after having studied prose passages. Participants took one or three immediate free-recall tests (on the entire passage), without feedback, or restudied the material the same number of times as the students who received tests. Specifically, one group was tested three times after having studied the passage once, and they recalled about 70% of the material on each test. Another group of participants was tested only once after having studied the passage three times, recalling 77% on the test. The last group had the highest exposure to the study material, studying the passage four times and were only tested on the final delay, 2 days later. Therefore, the participants in all three groups were exposed to the material four times, either via various study or test events. A final retention test was given after 5 minutes, 2 days, or one week later. Repeated studying improved recall relative to repeated testing only when the final test was given after 5 minutes. In contrast, when the final test was given 2 days or a week later, exactly the opposite performance emerged, recall was the highest in the conditions with the most repeated testing opportunities. The interesting finding in these experiments is that this effect occurred despite the fact that participants in the repeated study condition received much more exposure to the material. Roediger and Karpicke (2006) showed that repeated retrieval of the entire material greatly slows forgetting in long-term memory (for similar findings see: Wheeler, Ewers, & Buonanno, 2003; Karpicke & Roediger, 2008; Karpicke, 2009).

Several other experiments have established that retrieval practice produces a mnemonic boost relative to a restudy condition (e.g., Carpenter & DeLosh, 2005, 2006; Cull, 2000; Pyc & Rawson, 2007). Research suggests that, even though the use of different retrieval tasks may enhance learning based on different component processes, retrieval, no matter its nature (i.e., episodic or semantic) probably affects new learning based on the same general mechanism (Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Divis & Benjamin, 2014; Kornell & Vaughn, 2016; Finn, 2017). Though several explanations have been proposed to account for individual findings in the literature, to date no formal theory can independently account for the retrieval practice effect.

The most common theories which have been proposed thus far are summarised by Rickard and Pan (2018), and include: (1) Bjork's (1994) desirable difficulties model, suggesting that superior learning is achieved with more difficult retrieval, compared to both restudy or less difficult retrieval; (2) the distribution-based bifurcation model which proposes that testing with no feedback results in a bifurcation of the distribution of memory strength by response accuracy (Kornell, Bjork, & Garcia, 2011; Halamish & Bjork, 2011); (3) Carpenter's (2009) elaborative retrieval hypothesis which proposes that when retrieval is provided with feedback this creates additional associative paths between cue and response than does restudy; (4) the mediator effectiveness hypothesis by Pyc and Rawson (2010) which proposes that test trials are more likely to form cue-response mediators than restudy trials; (5) the episodic context theory which explains the retrieval practice effect through the differences in retrieval frequency and updating of the degree of episodic context (Karpicke, Lehman, & Aue, 2014) (6) Bouwmeester and Verkoeijen (2011) proposed the gist-trace processing account according to which memory is strengthened at the semantic level through testing while only memory for surface features is strengthened during restudy; and (7) the attenuated error correction theory which proposes that error correction processes in a feed forward two-layer neural network model explain the retrieval practice effect, when feedback is provided after testing (Mozer, Howe, & Pashler, 2004). Either most of the previously mentioned theoretical explanations are presented as speculations, or they only specifically apply to a subset of observed phenomena (Rowland, 2014). A detailed discussion on these is beyond the scope of the present paper (for a review see Rowland, 2014).

In contrast to the retrieval practice effect, which arises when an entire class of events (whole set of material) is retrieved, retrieval of a random selection of studied items at test can impede memory performance of the remaining items. Thus, in these instances the beneficial effect of retrieval often reverses into a detrimental effect (Slamecka, 1968; Roediger, 1973) which is usually termed retrieval induced forgetting and has proved to be remarkably robust (see meta-analysis by Murayama et al., 2014).

2.1.3. Retrieval of sub-parts of material.

Retrieval of a random selection of studied items at test has been shown to enhance their subsequent recall (Carpenter, 2012; Delaney, Verkoeijen, & Spirgel, 2010; Roediger & Butler, 2011; Roediger & Karpicke, 2006) at the expense of non-retrieved items (Anderson, 2003; Storm & Levy, 2012). Though several cognitive mechanisms have been attributed to this effect, to date no account has proved able to explain the full range of findings (Hulbert, Shivde, & Anderson, 2012; Storm & Levy, 2012; Verde, 2012; Jonker, Seli, & MacLeod, 2013; Raaijmakers & Jakab, 2013).

In apparent contradiction to the retrieval induced forgetting literature, more recent findings provide evidence that retrieval of sub-parts of material, can also enhance later recall of the non-retrieved subparts (so long as the two are related to each other; Chan, McDermott & Roediger, 2006). This phenomenon is termed retrieval induced facilitation. The literature investigating retrieval induced facilitation is thoroughly reviewed in a paper by Anderson (2003), while the relation between retrieval induced facilitation and retrieval induced forgetting is covered by Chan (2009). Briefly, two factors appear to determine whether initial testing facilitates or impairs later recall of non-tested material, the level of material integration and delay (Chan, 2009 but see e.g., Wallner & Bäuml, 2017; Bäuml, 2019; Wirth & Bäuml, 2020 for alternative accounts). Specifically, retrieval induced forgetting is eliminated if participants can integrate the studied material into a coherent and interconnected episode (e.g., Anderson, Green, & McCulloch, 2000; Bäuml & Hartinger, 2002; Migueles & Garcia-Bajos, 2007) and if there is a considerable delay interval (i.e., 24 h or more) between initial and final test (e.g., Bjork, Bjork, & MacLeod, 2006; Chan, McDermott & Roediger, 2006).

Chan, McDermott and Roediger (2006) suggested that the retrieval induced facilitation effect occurs because, during test, participants perform an active search of information related to the target memory. Chan (2009) proposes two assumptions to account for this explanation. Firstly, that the retrieval induced facilitation effect and the testing effect possibly share the same underlying mechanism. Specifically, a conscious and active retrieval process which enhances memory of the retrieved information, irrespective of whether this information was retrieved directly (as in whole list retrieval) or in a collateral fashion (as in retrieval of sub-parts of material). Secondly, he suggests that retrieval induced facilitation is more likely to be attributed to a controlled rather than an automatic process (i.e., spreading activation). As a result, as in the case of the directly tested facts in the testing effect, they should be strengthened over a long period of time (unlike semantic priming).

The previous sections have briefly reviewed studies demonstrating instances when testing benefits and hinders retention, discussing the distinction between the attempt to recall the whole set of studied material, and possible links to retrieval induced forgetting findings. One possibility is that both may be operating in the Roediger effect, but with participants in the Roediger effect simply showing a preponderance of the learning advantage over the effect of inhibition. The data that I will go on to discuss suggests a further possibility, potentially reflecting a third effect, that of implicit priming of existing representations. This new study investigated this priming hypothesis. If this effect is due to priming existing memory representations, then it should be preserved in groups that show a decline in learning ability but in whom there is intact priming.

There is substantial evidence that despite an age-related decline in learning ability and explicit memory, priming is preserved both in healthy ageing and amnesia (e.g., Camus et al., 2003; Lustig & Buckner, 2004; Bennett et al., 2006; Yano et al., 2008). Priming is typically described as an activation of a representation in memory which facilitates performance on tasks involving that representation (Java & Gardiner, 1991). Tasks employed in measuring priming differ from more conventional memory tests in that they do not require explicit recollection. Priming is said to manifest across a range of modalities, including verbal memory, where it provides a way in which even densely amnesic patients can appear to have normal memory (Baddeley, Eysenck & Anderson, 2015, p. 459).

Shimamura (1986) has proposed that once encoding has been successful during study, the priming of new associations reflects a process that is preserved even in amnesic patients. Specifically, while declarative memory is required to encode new information, once this is successful, priming proceeds automatically at retrieval (Shimamura, 1986). Data published by Stamate, Logie, Baddeley, and Della Sala (2020), using the same method as in the present study (namely repeated retrieval of subparts of material), also seems to support this idea (these data are presented in Chapter 3). Those results showed that Alzheimer's patients, after being taught to a specific learning criterion (70% correct), benefited from repeated retrieval (by sampling different features from each narrative on every test session/delay, with no feedback) to the same extent as healthy controls. Both the fact that Alzheimer's patients exhibited learning deficits, and because the design minimised relearning, suggests that the act of repetition served to strengthen existing representations and was likely dependent on priming.

The present experiment aimed to test if the retrieval induced facilitation effect is dependent on priming or re-learning, by comparing healthy older with healthy younger participants. If priming is key, then the age-related decline in explicit learning should result in older participants taking longer to learn than younger, but intact effects of priming in the older group should result in little or no difference between groups in forgetting rates when tested repeatedly with different feature probes at different intervals after learning. The current test was specifically designed never to probe the same features of the encoded material. Instead, it samples different subparts of the material (learned only at encoding) on each testing occasion, without ever re-exposing participants to the entire material or providing feedback in order to minimise re-learning. The selection of material was based on The Crimes Test (Baddeley, Rawlings, & Hayes, 2014), which is atypical of normal prose in using a matrix structure comprising four crimes and five features which can be presented in different combinations. This test was created as a solution to the need for a reliable measure of long-term forgetting in clinical settings (e.g., with amnesic patients). It consists of four short stories, each based on a crime, comprising five key features: the crime, the criminal, the location, the age/sex of the victim and the nationality of the victim (e.g., ‘A young Chinese woman had arranged to meet her sister at morning service. Just as she was about to enter the cathedral she noticed a child who seemed to be begging. He suddenly snatched her handbag and ran off.’). The test does not demand excessive (initial) learning time and allows for different subsamples of questions to be tested via cued recall (e.g. ‘Which crime occurred against the young woman?’) after a range of delays. Therefore, the test has the capacity to provide a reliable measure across repeated tests over separate test sessions. For the experiments reported in this thesis, this principle was extended to materials structured as normal prose presented as short fables (see full description of materials in section 2.2.2).

The goal was to answer the following questions:

1. Do we observe the age effect on learning shown in previous studies?
2. Does the selective feature sampling method at various delays reduce forgetting?
3. If the above is observed does it reduce any differences in forgetting rates between older and younger participants?
4. Is there evidence of accelerated forgetting rates in the older group?

The generality of the results from Experiment 1 was assessed in Experiment 2 using a design that disrupts the overall theme of each passage by scrambling the order of the constituent sentences in each prose passage. If the priming advantage remains, it will suggest that it continues to operate at the sentence level, where different features of a given sentence are probed after different delays.

2.2. Method.

2.2.1. Participants.

A total of 60 young participants were recruited mainly from Carol Davila University of Medicine and Pharmacy of Bucharest and Transilvania, University of Brașov, Romania, and a few from among friends and acquaintances. The 60 older participants were all healthy, community-dwelling adults recruited from the Brașov City Hall Seniors’ Club, through GPs, and a few from among friends and acquaintances. Participants’ written consent and demographic information concerning education, self-report of health and medication, and past or present medical conditions, were obtained before starting the experiment. There were no statistically significant differences in the number of years of education between younger

(M= 13.96, SD= 1.12) and older participants (M= 13.36, SD= 2.47) [F (1, 119) = 732, p= .394].

The 60 younger (without retrieval practice condition age: M= 23.6, SD= 1.7; retrieval practice condition: M= 23.1, SD= 2.3) and 60 older (without retrieval practice age: M= 69, SD= 6.7; retrieval practice condition: M= 68, SD= 8.2) healthy participants were assigned at random to either a condition with retrieval practice or a condition without retrieval practice (see Figure 2.1), adding up to 60 participants in each delay condition. For the encoding phase participants were tested in person, for the remaining test delays participants were tested via telephone.

2.2.2. Material.

The material was inspired by some unfamiliar fables by Aesop (Winter, 1919). To make the material more appropriate for older participants, the animal characters from the original fables were changed into human characters (i.e., a French woman instead of a bear). 4 fables were created, each was 4 sentences long and consisted of 13 combined features (e.g., 'A French woman used to take long walks up on a hill and came across a swarm of bees. One cloudy afternoon the young woman approached the swarm when two bees flew out and stung her.'; features: gender-nationality, i.e., 'What was the sex of the French person?'; action-animal, i.e., 'What animal/animals stung a character?'; nationality-location, i.e., 'What was the nationality of the character from the hill?'— see full material in the Appendices). This generated 52 questions, which were split between 4 subsets, each question in the subsets probed a single sentence from the 4 fables, without ever probing the same sentence within the same subset (see Probing Scheme in the Appendices). The material and question sets closely followed the Crimes Test structure (Baddeley, Rawlings, & Hayes, 2014).

2.2.3. Design.

The experiment employed a between-subjects design. Participants in the retrieval practice condition were tested at 4 delays: post-encoding retrieval, 1 day, one week and 1 month. At each delay the test involved a different subset of questions. Participants in the condition without retrieval practice were only tested at 2 delays: post-encoding and after one month.

All testing was conducted in Romanian (my native language and that of all participants). The material (both the fables and the material with the disrupted theme), was initially devised and piloted in English and subsequently translated into Romanian.

During the encoding phase, all participants were presented with the four fables, these were read out by the experimenter (the author of this thesis) at a slow and clear pace (2s pause between each sentence and a 5s pause between each fable). To minimise any recency effects, each presentation phase was followed by a 1-minute filler task that involved creating as many words as possible from the Romanian word 'hippopotam' (see Baddeley, Allen, Atkinson & Kemp, 2019). Participants then took the initial post-encoding cued recall test on one of the four subsets of the material. This was self-paced. If participants scored less than 70% correct (9 out of 13 questions), the 4 fables were presented again (in a different order), they took the 1-minute filler task again (creating new words) and were then retested. This process was repeated until the participant reached the set 70% criterion. The encoding phase and initial test were conducted face to face while all other tests were conducted by telephone. This type of testing has been validated in previous studies (e.g., Geffen et al., 1997; Baddeley, Rawlings, & Hayes, 2014) and used successfully in subsequent studies using procedures similar to those employed in the current experiment (Baddeley, Allen, Atkinson & Kemp, 2019).

2.2.4. Power analysis.

Statistical power analysis was performed to estimate the sample-size needed, to examine the effects of the different testing condition (without retrieval practice; with retrieval practice) between younger and older participants and any possible interactions with recall performance. With an alpha of 0.01 and power = 0.95, the estimated sample size needed for an effect size of d = 0.20 is N= 100 for a within-between groups interaction. The current sample size was slightly larger, final sample size of N = 120.

2.3. Results.

Mean recall scores on post encoding and 1 month by younger and older groups for each testing condition are displayed in Figure 2.1. Individual performance means for each group are displayed in Figure 2.2a, b, c & d in Supplementary Material).

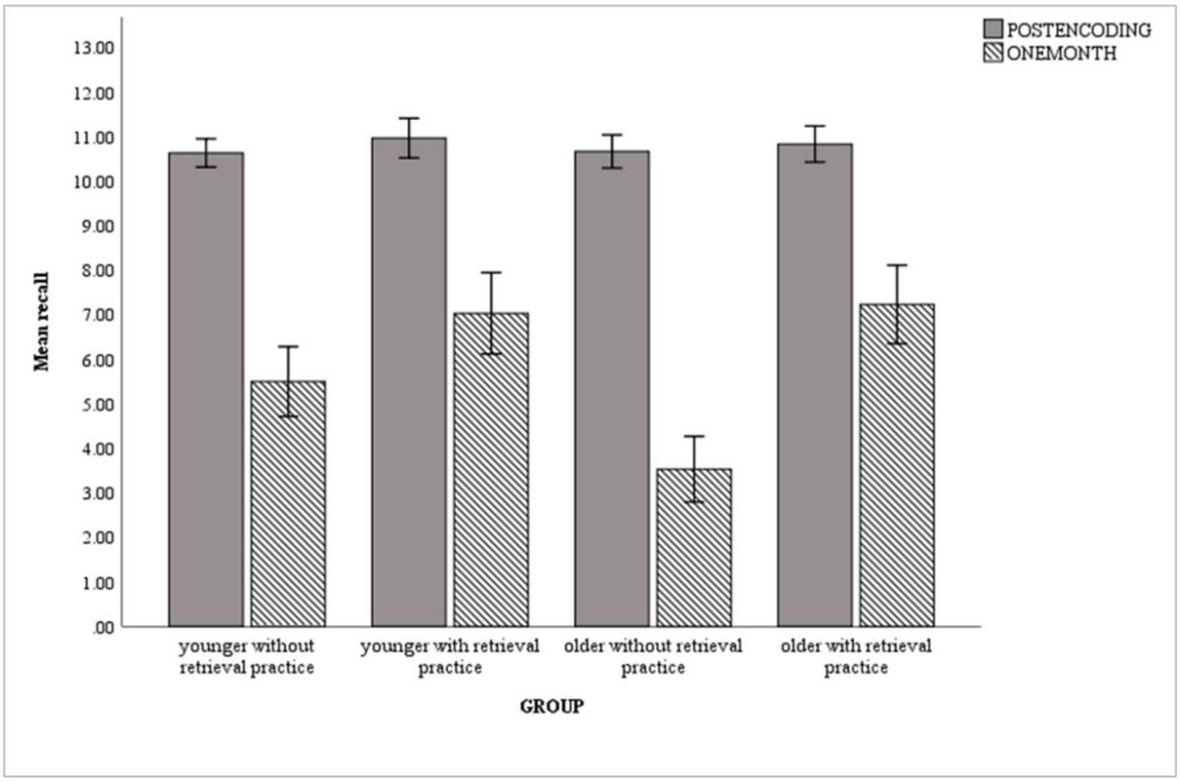


Figure 2.1 Mean recall scores at post encoding and 1-month tests by younger and older groups in the condition with retrieval practice and in the condition without retrieval practice.

Data were analysed employing a null-hypothesis significance testing (NHST) approach - analyses of variance (ANOVAs) and Bayesian hypothesis testing approach -Bayesian ANOVA.

To assess the change in mean recall performance from the post encoding assessment to the 1-month assessment, as a function of testing condition and age, a 2 by 2 by 2 mixed analysis of variance (ANOVA) with $\alpha = 0.05$, with delay (post encoding, 1 month) as a within-subjects factor and condition (without retrieval practice, with retrieval practice), and age group (younger, older) as between-subjects' factors was conducted. The Bayes Factors (BF) in favour of the alternative hypothesis (i.e., BF_{10}) for the main effect models including delay, condition or age were calculated by comparing the alternative models against the null model. The BF_{10} for the two-way interaction models was calculated as the ratio BFInteraction Model (the model with both main effects and the interaction)/ BFMain Effects Model (the model with both main effects). The BF_{10} for the three-way interaction model was calculated as the ratio BFThree-way Interaction Model (i.e., the model with both main effects, all two-way interactions, and the three-way interaction)/ BFMain Effects and Two-way Interactions Model (the model including the 3 main effects and the two-way interactions) (Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2017). A BF between 1-3 is considered as weak evidence, between 3-20 was considered as positive evidence, between 20-150 as strong evidence, and larger than 150 as very strong evidence.

The analysis showed that performance was significantly different at the post encoding assessment compared to the 1-month assessment (significant main effect of delay [F (1, 116) = 596.285, p<.001, ηp2= .837, BF10= 4.1*1050]) with lower performance at 1-month assessment. ANOVA results showed that the effect of delay on recall performance was larger overall for the older groups than the younger groups (a significant delay by age-group interaction [F (1, 116) = 4.225, p= .042, ηp2= .035]), and the effect of delay on recall performance was larger in the condition without retrieval practice than in the condition with retrieval practice (a significant delay by condition interaction [F (1, 116) = 37.077, p<.001, ηp2= .227]). The Bayesian analysis, however, provided no evidence for the delay by age-group interaction, but very strong support for the model with the delay by condition interaction (BF_{10} =1*106). The three-way delay by condition by age-group interaction was significant [F (1, 116) = 8.281, p= .005, ηp2= .067]. A Bayesian mixed factor ANOVA provided positive evidence in favour of the alternative model, that recall performance depends on the three-way interaction between delay, testing condition and age-group (BF_{10} =9.65).

Due to the significant three-way interaction, as well as the a priori interest in whether retrieval practice mediates the decline in recall performance, particularly in older adults, separate 2 by 2 (delay by condition) mixed ANOVAs with pairwise comparisons (Bonferroni corrected) and Bayesian mixed factor ANOVAs at each delay, were carried out for each age group. Finally, age-related changes in recall performance were analysed, by conducting separate 2 by 2 (delay by age-group) mixed ANOVAs and Bayesian mixed factor ANOVAs in each testing condition (with retrieval practice vs. without retrieval practice).

2.3.1. The effect of testing condition: mixed ANOVAs comparing younger groups' performance revealed significant main effects of delay [F (1, 58) = 265.537, p<.001, ηp2= .821], and testing condition [F (1,58) = 6.656, p= .012, ηp2= .103]. The delay by testing condition interaction was significant [F (1, 58) = 4.651, p= .035, ηp2= .074] and was explained by the fact that the group in the testing condition with retrieval practice performed significantly better at 1-monthcompared to the group tested in the condition without retrieval practice (MD= 1.533, p= .012) while performance as at post encoding assessment was similar for between groups (MD= .333, p= .217). Results of the Bayesian ANOVAs provided weak evidence for the alternative model with the delay by condition interaction (BF_{10} =1.9) and positive support (BF_{10} =4.2) for the main effect of condition on recall performance at the 1-month assessment.

An identical pattern of results was seen when comparing the older groups' performance: there was a significant main effect of delay [F (1, 58) = 331.149, p<.001, ηp2= .851] and testing condition: [F (1, 58) = 35.886, p<.001, ηp2= .376] qualified by a significant delay by condition interaction [F (1,58) = 34.904, p<.001, ηp2= .382]. The two older groups had similar performance on the post encoding assessment (MD= 167, p= .538) but the group in the testing condition with retrieval practice performed significantly better at 1-monthcompared to the group tested in the condition without retrieval practice (MD= 3.700, p<.001). Results of the Bayesian ANOVAs provided very strong evidence for the alternative model with the delay by condition interaction (BF_{10} =887,640) and for the main effect of condition on recall performance at the 1-month assessment (BF10=643,286). Simply put, the decline in recall performance over the 1-month interval was faster in the condition without retrieval practice irrespective of age-group (Fig. 2.2).

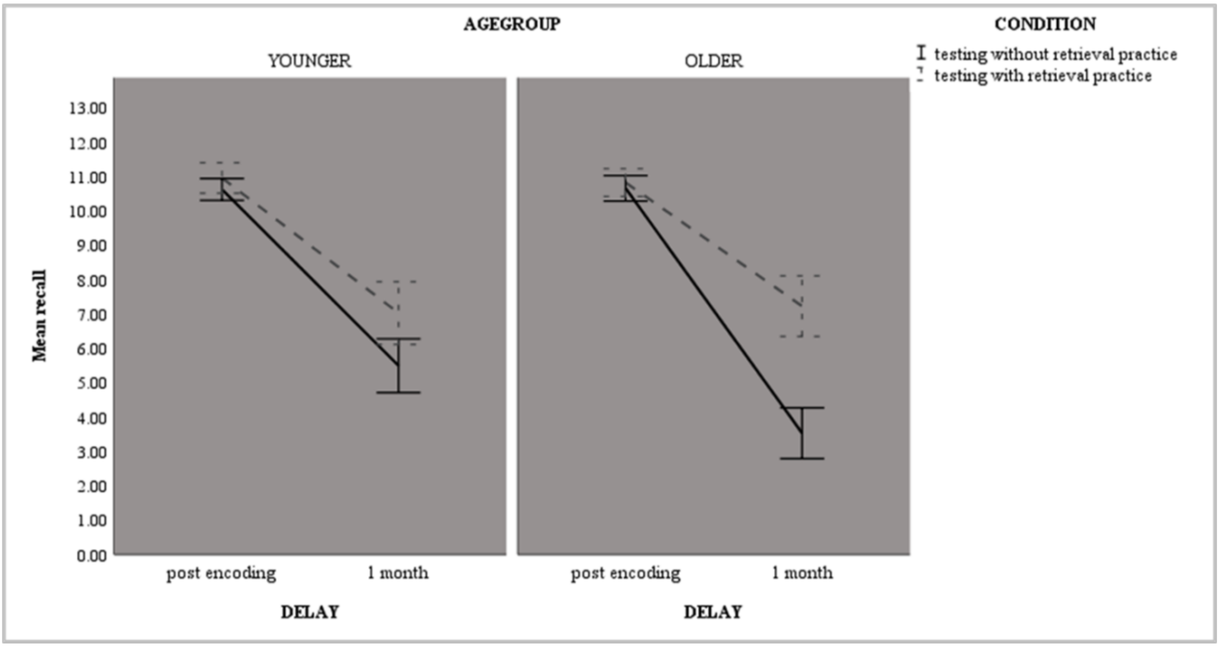


Figure 2.2 Mean recall performance of younger and older groups at post-encoding and 1-month intervals in the condition without retrieval practice and the condition with retrieval practice using material with an integrated narrative.

2.3.2. The effect of age: mixed ANOVA comparing younger and older adults' performance in the testing condition without retrieval practice provide strong support for AFL in older adults. Results showed significant main effects of delay [F (1, 58) 453.080, p<.001, ηp2= .887] and age-group [F (1,58) = 11.126, p<.001, ηp2= .161]. These significant main effects reveal that both age groups' performance declined significantly from the post encoding assessment to the 1-month assessment and the fact that the older groups' overall performance was lower still compared to that of the younger group. Importantly there was a significant interaction between delay and age group [F (1, 58) = 12.044, p<.001, ηp2= .172] explained by significant differences in performance between groups at the 1-month assessment only (MD= 1.967, p<.001,). Results of the Bayesian ANOVAs provided strong support for the delay by age group interaction (BF_{10} =97.22) and for the main effect of age group on recall performance at 1-month assessment (BF_{10} = 63.46).

In the condition with retrieval practice no significant effect age group was found [F (1,58) = .077, p= .933, ηp2<.001] and no significant delay by age group interaction [F (1, 58) = .341, p= .561, ηp2= .006]. Therefore, in the testing condition with retrieval practice recall performance was similar between younger and older groups both on post-encoding retrieval assessment (MD= .133, p= .264) as well as at 1-month (MD= -.200, p= .748 - see Fig 2.3).

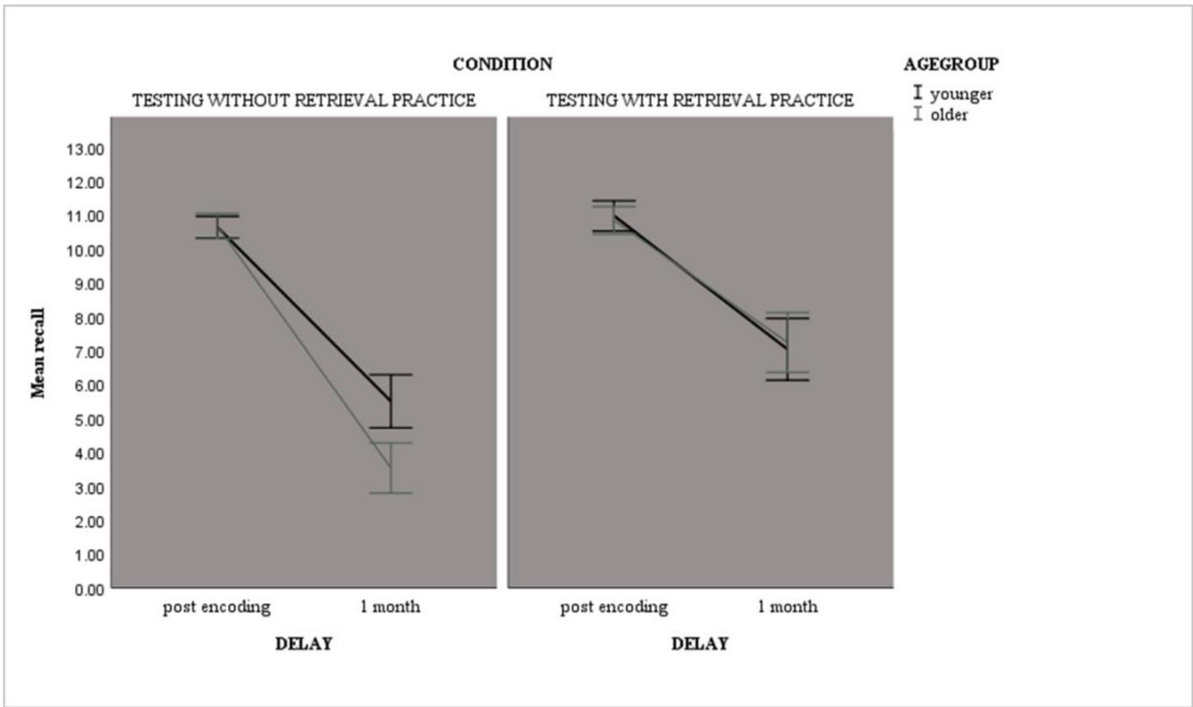


Figure 2.3 Mean recall performance of younger and older groups at post-encoding and 1-month intervals in the condition without retrieval practice and the condition with retrieval practice using material with an integrated narrative.

A significant encoding deficit relates to the number of trials necessary to reach the 70% criterion at encoding. A chi-square test of independence showed a statistically significant effect of age-group on the number of trials [X2 (1, N = 120) = 15.4 φ=.36, p < .001] with more older participants (n= 29) than younger participant (n=9) requiring more than one trial to reach criterion.

Analyses were also conducted to explore whether mean recall performance significantly differed across the 4 delays in the condition with retrieval practice. A 2 (age group: younger, older) by 4 (delay: post encoding, 1 day, one week, 1 month) repeated measures ANOVA revealed a main effect of delay [F (3, 118) = 164.047, p <.001, ηp2= .595], but no significant effect of age-group [F (1, 118) = .056, p = .814, ηp2= .018] and no significant delay by age-group interaction [F (3, 118) = 1.573, p= .199, ηp2= .023]. Further Bonferroni corrected Pairwise Comparisons showed that there were no differences in performance between younger and older adults on any of the 4 assessments. There was a statistically significant difference between mean score at post encoding assessment test compared to mean scores at all 3 delay intervals in both age-groups. The only significant decline in recall performance between two consecutive assessments was between the post encoding test and the 1-day test for both younger and older participants (younger: MD= 3.150, SDE= .216, $p<.001$; older: MD= 2.433. SDE= .392, $p<.001$).

2.3.1. Summary of results.

These results suggest that testing condition does have an effect on forgetting rate. Specifically, the results show that repeatedly retrieving sub-parts of material benefits retention and attenuates the forgetting rate in both younger and older groups.

Age related differences in recall performance over the 1-month interval were only found in the condition without retrieval practice. While in the condition with retrieval practice, older and younger participants showed a similar decline in performance.

Results in the testing condition with retrieval practice showed that the most significant decline in performance occurred between post encoding assessment and one day assessment for both younger and older participants.

2.4. Discussion.

Experiment 1 showed that forgetting rates were reduced over the course of one month, for both healthy older and younger adults, as a result of repeatedly retrieving sub-parts of material, in contrast to larger drops in performance seen with no repeated testing. This supports the idea that this sampling method delays forgetting. Furthermore, the pattern of results indicates that repeatedly retrieving sub-parts of material delays forgetting as a result of priming rather than relearning. The older participants in this study did take longer to learn, showing the usual age effect on learning. After having equated baseline performance, they showed reduced and equal forgetting rates comparable to that of the younger participants when tested repeatedly. This memory advantage most likely arises due to the fact that the integrated material contains multiple features, therefore when it is probed several times at different delays, even though this is done with different questions (features), participants will likely retrieve the entire encoded episode.

Lastly the performance patterns from the condition with no repeated testing seems to provide evidence in favour of the literature claiming accelerated forgetting rates in the older groups.

When memory was not refreshed (probed with the intervening tests) over the course of one month, the older group showed accelerated forgetting rates compared to the young.

Although these results provide evidence for an impact of priming from testing different sub parts of a prose passage, it is not clear if this arose because participants retained an integrated representation of the sequence of events in each prose passage. For example, presenting one fragment of text might have primed retrieval of the complete story line for the prose from which that fragment was selected. It remains an open question as to whether the priming might also appear at the sentence level without encoding an integrated passage of prose. In Experiment 2, this latter possibility was investigated by using material in which the overall theme and sequence of events in each passage was disrupted by scrambling the order of the constituent sentences. If the effect of repeated partial testing remains it will suggest that the priming effect can operate at the sentence level, when different features of a given sentence are probed after different delays. This issue was investigated using the same design (without retrieval practice condition vs. a retrieval practice condition) and exposing a different sample of participants to the new material (full description in the section 2.6.2).

Experiment 2: Sampled testing of material with a lower level of integration. Does the priming advantage remain?

2.5. Introduction.

The generality of order of the results of Experiment 1 was assessed using a design that disrupts the overall theme of each passage by scrambling the constituent sentences. If the priming advantage remains, it will suggest that it continues to operate at the sentence level, where different features of a given sentence are probed after different delays.

2.6. Method.

2.6.1. Participants.

60 younger (without retrieval practice condition: M= 24, SD= 3.4; retrieval practice condition: M= 24.6, SD= 3.6) and 60 older (without retrieval practice condition: M= 68.1, SD= 5.9; retrieval practice condition: M= 68.4, SD= 5.9) were recruited from the same

sources as Experiment 1. None took part in Experiment 1. They were assigned at random to either the retrieval practice condition or the without retrieval practice condition (Figure 2.2), adding up to 60 participants in each delay condition. For the encoding phase participants were tested in person, either individually or in groups of a maximum of five, for the remaining test delays participants were tested via telephone. There were no statistically significant differences in the number of years of education between younger (M= 15.05, SD= 1.71) and older participants (M= 14.73, SD= 1.88) [F (1, 119) = 933, p= .336].

2.6.2. Material.

For the material with the disrupted narrative the aim was to keep it as similar as possible to the integrated material in terms retrieval difficulty, while attempting to disrupt the creation of narrative structure. This was achieved by scrambling the sentences across the 4 fables and deriving 16 individual sentences (i.e., ‘The moral of the story with the wise frog is about pleasure: We should not take pleasure at the expense of others; A Spanish woman lived quietly in her white house in a small village with a lot of dogs.’- see Appendices). Using the same sentences from the integrated material also allowed for the same question sets to be used with this material.

2.6.3. Design.

The experiment was conducted exactly the same as Experiment 1 but with the new material (disrupted narrative).

2.6.4. Power analysis.

For the data in Experiment 2, the same sample-size estimate as for Experiment 1 was used.

2.7. Results.

Mean recall scores on all testing sessions by younger and older groups for each testing condition are displayed in Figure 2.4. Individual performance means for each group are displayed in Figure 2.2.a, b, c & d in Appendices).

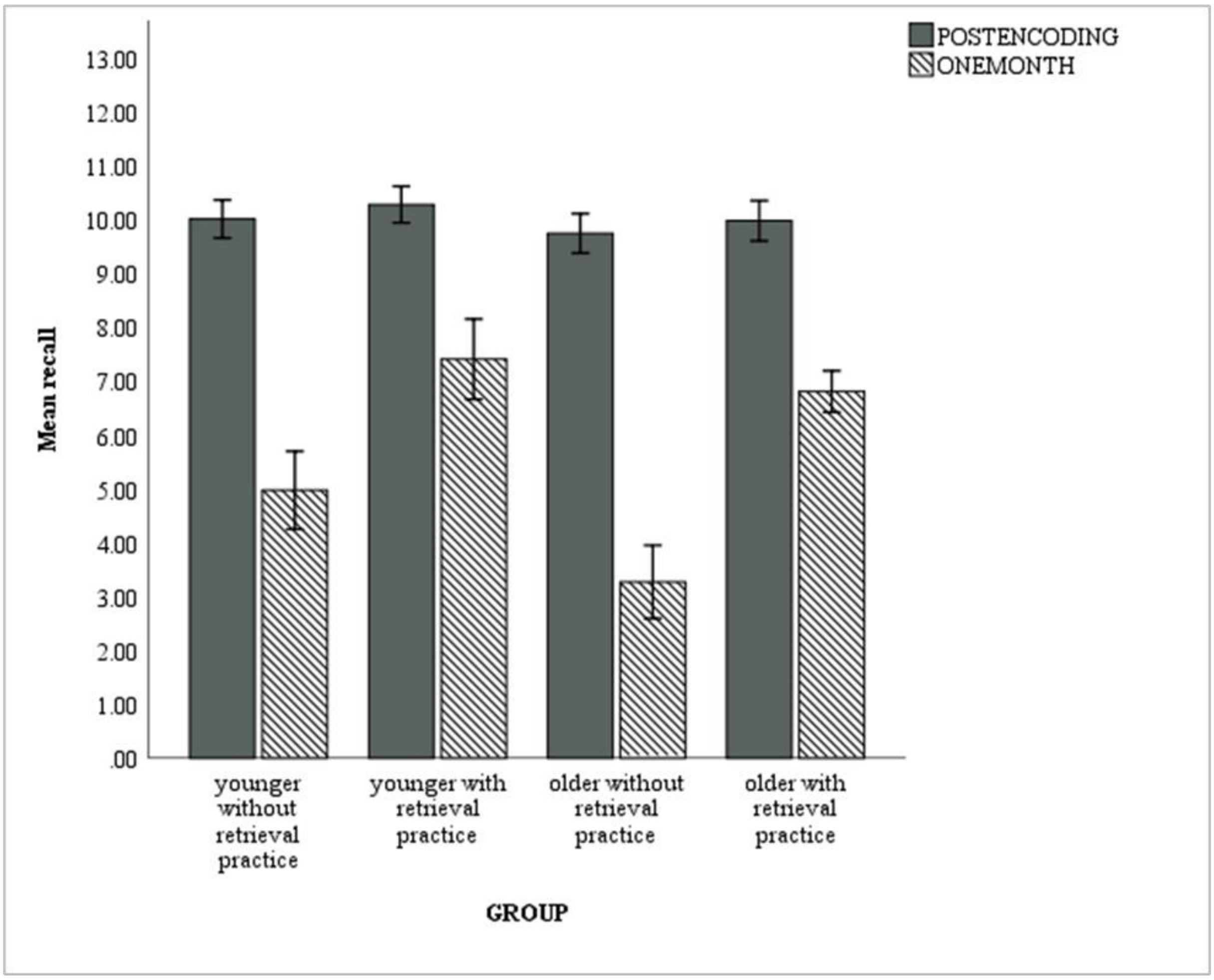


Figure 2.4. Mean recall performance of the older and younger groups at different delay intervals in the condition without retrieval practice and the condition with retrieval practice using the material with a disrupted narrative.

To assess the change in mean recall performance from post encoding assessment to 1-month assessment as a function of testing condition and age a 2 by 2 by 2 mixed analysis of variance (ANOVA) and Bayesian mixed factors ANOVA, with $\alpha = 0.05$, with delay (post encoding, 1 month) as a within-subjects factor and condition (without retrieval practice, with retrieval practice), and age group (younger, older) as between-subjects factors was conducted. Results showed that mean recall performance was significantly different at post encoding assessment compared to 1 month assessment (significant main effect of delay [F (1, 116)= 689.165, $p<.001$, $\eta p2= .856$, $BF_{10}=5{,}9*10^{49}$] with lower performance at 1 month assessment. There was an overall difference in recall performance between age groups (significant main effect of age-group: [F (1, 116) = 13.566, $p<.001$, $\eta p2= .105$]). The overall difference between testing conditions (without vs. with retrieval practice) was significant [$F(1, 116) = 69.03$, $p<.001$, $\eta p2= .373$].

The effect of delay on recall performance was larger overall for the older groups compared to the younger groups (a significant delay by age-group interaction [$F(1, 116) = 6.736$, $p= .011$, $\eta p2= .055$]), and the effect of delay on recall performance was larger in the condition without retrieval practice compared to the condition with retrieval practice (a significant delay by condition interaction [$F(1, 116) = 66.997$, $p<.001$, $\eta p2= .366$]). The Bayesian analysis, however, provided very weak support for the model including the delay by age-group interaction ($BF_{10}=1.3$) but very strong support for the delay by condition interaction ($BF_{10}=520,000$).

The three way delay by condition by age-group interaction was not significant [$F(1, 116) = 2.880$, $p= .092$, $\eta p2= .024$, $BF_{10}=.446$].

To further investigate the effect of condition 2 by 2 (delay by condition) mixed ANOVAs with pairwise comparisons each delay (Bonferroni corrected) were ran, separately for each age group, as well as Bayesian mixed ANOVAs. Finally, age-related changes in recall performance were analysed by running separate 2 by 2 (delay by age-group) mixed ANOVAs and Bayesian mixed ANOVAs in each testing condition (with retrieval practice vs. without retrieval practice).

2.7.1. The effect of testing condition (with retrieval practice vs. without retrieval practice)

The analysis of the younger groups performance revealed a significant main effect of delay [$F(1, 58) = 211.738$, $p<.001$, $\eta p2= .785$, $BF_{10}=1*10^{22}$], and testing condition [$F(1,58) = 21.748$, $p<.001$, $\eta p2= .273$, $BF_{10}=4.49$] and a significant interaction between delay and testing condition (younger: [$F(1, 58) = 15.927$, $p<.001$, $\eta p2= .215$]. Pairwise comparisons revealed that the two younger groups had similar performance at post encoding assessment ($MD= .267$, $p= .239$) but the group in the testing condition with retrieval practice performed significantly better at 1-monthcompared to the group tested in the condition without retrieval practice ($MD= 2.433$, $p<.001$). Bayesian analysis provided very strong support for the model including the delay by condition interaction ($BF_{10}= 409$) and for the main effect of condition on recall performance at the 1-month assessment ($BF_{10}=1455$).

An identical pattern of results was seen when comparing the older groups performance: there was a significant main effect of delay [$F (1, 58) = 613.847$, $p<.001$, $\eta p2= .915$, $BF10=4*10^{28}$] and testing condition: [$F (1,58) = 52.443$, $p<.001$, $\eta p2= .475$, $BF_{10}=20$] qualified by a significant delay by condition interaction [$F (1, 58) = 71.975$, $p<.001$, $\eta p2= .554$]. As shown in Figure 2.5, the two older groups had similar performance at post encoding assessment (MD= .233, p= .365) but the group in the testing condition with retrieval practice performed significantly better at 1-monthcompared to the group tested in the condition without retrieval practice (MD= 3.533, p<.001). Bayesian analysis provided very strong support for the model including the delay by condition interaction (BF_{10}=937,500) and for the main effect of condition on recall performance at 1-month assessment ($BF_{10}=1.04*10^{10}$).

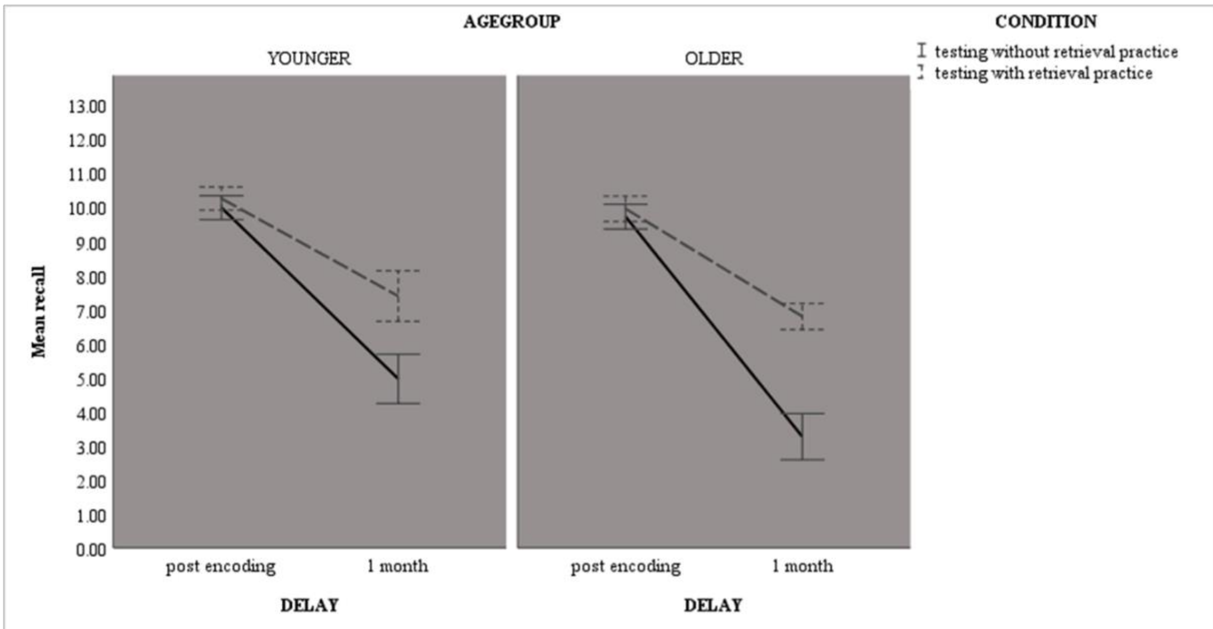


Figure 2.5. Mean recall performance of younger and older groups at post-encoding and 1-month intervals in the condition without retrieval practice and the condition with retrieval practice using material with a disrupted narrative.

Similarly, to the integrated material, there a was a significant drop in mean scores at 1-month test compared to the post-encoding retrieval test but participants who were tested repeatedly across the 1- month interval had a lower forgetting rate compared to participants who were tested post-encoding and at 1-month only. These results suggest that testing condition does have an effect on forgetting rate. Specifically, the results show that repeatedly retrieving sub-parts of material over 1-month attenuates the forgetting rate in both younger and older groups.

The effects of age: The two-way mixed ANOVA in the testing condition without retrieval practice provide support for AFL in older adults. There were significant main effects of delay [$F (1, 58) = 579.829$, $p<.001$, $\eta p2= .909$] and age group [$F (1,58) = 10.550$, $p= .002$, $\eta p2= .154$] and a significant interaction between delay and age group [$F (1, 58) = 9.007$, $p= .004$, $\eta p2= .134$]. Whereas younger and older participants had similar mean recall scores at the post-encoding retrieval assessment (MD= .267, $p= .288$), at the 1-month assessment younger participants had significantly higher mean recall scores compared to older participants (MD= 1.700, $p= .001$) showing a steeper decline in performance for older adults. The results from the Bayesian analysis provided positive support for the models with the delay by age-group interaction (BF_{10}=11.66) and strong support for the main effect of age-group (BF_{10}=34) on recall performance at 1-month assessment.

In the condition with retrieval practice no significant effect age group [$F (1,58) = 3.387$, $p= .071$, $\eta p2. = .055$] and no significant delay by age group interaction [$F (1, 58) = .413$, $p= .523$, $\eta p2= .007$] was found. Therefore, as seen in Figure 2.6, in the testing condition with retrieval practice recall performance was similar between younger and older participant groups at the post-encoding retrieval assessment (MD= .300, $p= .288$) as well as at 1-month (MD= -.600, $p= .148$).

The effect of age-group: A significant effect of age relates to the number of trials necessary to reach the 70% criterion at encoding. A chi-square test of independence showed a statistically significant effect of age-group on the number of trials ($X2 (1, N = 120) = 7.12$, $\varphi=.25$, $p = 014$] with more older participants (n= 57) than younger participant (n=47) requiring more than one trial to reach criterion.

Therefore, age related differences in recall performance over the 1-month interval were only found in the condition without retrieval practice. In the condition with retrieval practice, mean recall scores at 1-monthwere similar for older and younger participants. These results suggest that both younger and older participants benefited from repeated retrieval by using different sub parts of sentences at each retrieval delay.

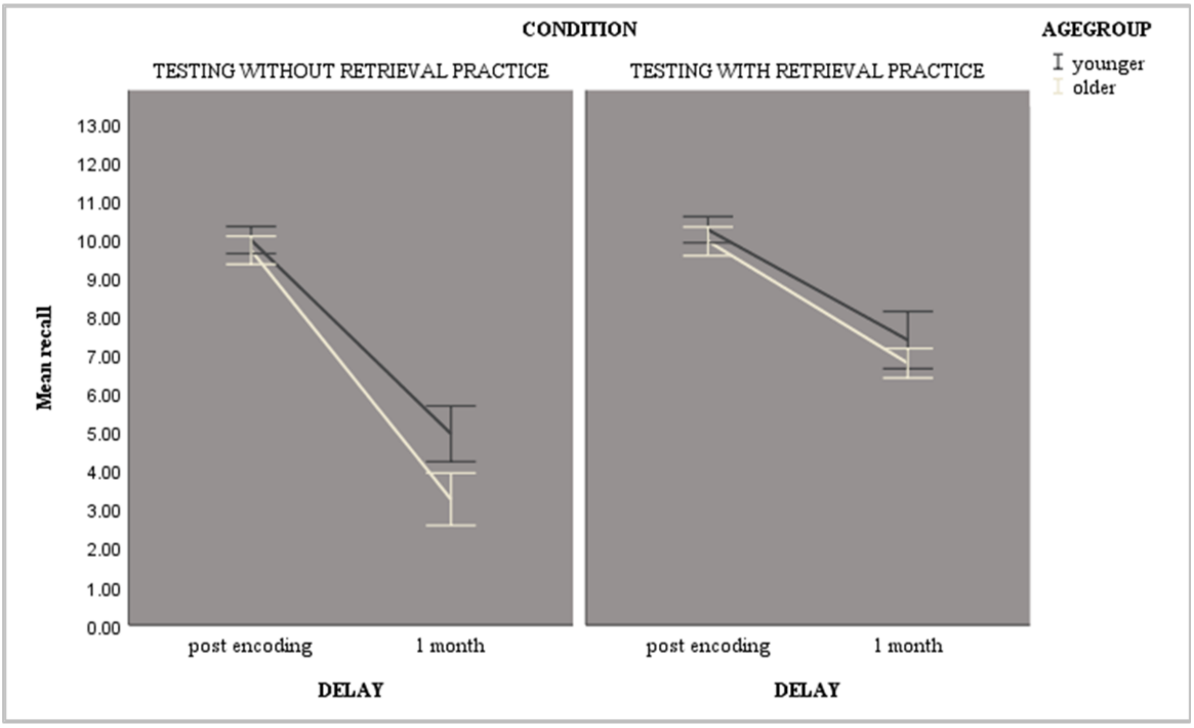


Figure 2.6. Mean recall performance of younger and older groups at post-encoding and 1-month intervals in the condition without retrieval practice and the condition with retrieval practice using material with a disrupted narrative.

Analyses were also conducted to explore whether mean recall performance significantly differed across the 4 delays in the condition with retrieval practice. A 2 (age group, younger, older) by 4 (delay: post encoding, 1 day, one week, 1 month) repeated measures ANOVA revealed a main effect of delay [F (3, 174) = 83.044, p<.001, ηp2= .589], no significant main effect of age [F (1,58) = 1.132, p= .314, ηp2= .017] and no significant delay by age by group interaction [F (3, 174) = .750, p= .390, ηp2= .020]. Further Bonferroni corrected Pairwise Comparisons showed that there were no differences in performance between younger and older adults on any of the 4 assessments. There was a statistically significant difference between mean score at post encoding assessment test compared to mean scores at all 3 delay intervals in both age-groups. The only significant decline in recall performance between two consecutive assessments was between post encoding and 1-day testing in both younger and older participants (younger: MD= 3.133, SDE= .345, p<.001; older: MD= 2.667, SDE= .227, p<.001).

2.7.2. General comparisons between recall performance with the integrated material and recall performance with the material with a disrupted narrative.

As opposed to the integrated material, the material with a disrupted narrative used in this experiment was more difficult to encode. Only a few participants, 21.7% (n= 13) of the younger and 5% (n= 3) of the older, reached criterion on the first trial. The majority needed two trials to achieve criterion: 70% (n= 42) of the younger and 90% (n= 54) of the older participants. The remaining participants took three trials to achieve criterion (younger: n= 5; older: n= 3). A regression model was run to evaluate the predictive power of the number of trials at encoding on forgetting at one month, the effect was not significant [F (1,119) = .010, p= .922].

To eliminate the possibility that the similar patterns in forgetting between the two experiments were not just an artefact of the adjusted initial performance, and to see the degree to which this was an effect of type of material (integrated vs disrupted narrative material) several statistical analyses were run.

A significant effect of material relates to the number of trials necessary to reach the 70% criterion at encoding. A chi-square test of independence showed a statistically significant effect for type of material on the number of trials [X2 (1, N = 240) = 75.12, φ=.56, $p < .001$] with more participants requiring more than one trial with the disrupted narrative material (n= 104) compared to integrated material (n= 38).

A Mixed ANOVA showed no significant material by age-group interaction [F (1, 232) = 1.080, p= .300], no significant material by condition interaction [F (1, 232) = .682, p= .410] no significant material by age-group interaction [F (1, 232) = .682, p= .410], no significant material by age-group by condition interaction [F (1-232) = .972, p= .325], but a significant delay by material interaction [F (1, 232) = .682, p<.001, $\eta p2$= .020].

The significant delay by material interaction effects were further investigated by conducting tests of simple main effects for each testing session to evaluate how the level of material integration may differentially impact mean recall scores. The only significant effect of material [F (1, 239) = 36.830, p<.001] was found at on post-encoding retrieval, with higher mean scores for the integrated material (M= 10.775, SD= 1.062) compared to material with a disrupted narrative (M= 9.917, SD= 1.072).

2.8. Discussion.

The pattern of results from this experiment is consistent with that in Experiment 1. The scrambled sentence order in the material used here led to more difficulty in encoding for both groups compared with Experiment 1, and this difficulty in learning was greater for older participants as found in previous studies. Crucially, repeatedly retrieving different sub-parts of material again delayed forgetting equally for both groups. Though the narrative structure of the material across the whole prose passage was disrupted, the individual sentences still contained multiple features, therefore they are in a way similar to the integrated material (and The Crimes test). Consequently, when the same sentence is probed several times at different delays, even though they are different probes (e.g. weather on one test, bees on another), it is likely that participants retrieve the whole of the episode within that sentence (e.g. the young woman, the cloudy afternoon, the swarm of bees and the two that stung her). This pattern of results would indicate priming occurring at the sentence level, probably in addition to priming at the story level seen in the first experiment.

2.9. General Discussion Experiments 1 and 2.

Both Experiments 1 and 2 showed that testing subparts of prose material on multiple occasions, delays forgetting of that material. The overall pattern was that of well-maintained performance across a 1-monthdelay in the condition with retrieval practice, in contrast to a marked loss in the condition without retrieval practice, with both kinds of material.

Because this effect was equivalent between younger and older individuals, it supports the idea that the advantage emerges because testing primes information which was previously stored during encoding. While older individuals are typically outperformed by the younger on tasks that depend on declarative memory, several studies have shown that priming is a process that is intact in old age, once information has been successfully encoded (e.g. Shimamura, 1986; Yano et al., 2008). The pattern of results in the current experiments, both from the ones described here and one using a similar methodology in a population with mild Alzheimer's disease (Stamate, Logie, Baddeley, & Della Sala, 2020; described in Chapter 3), would seem to fit with the explanation that priming is intact in populations with learning deficits and that it is the underlying mechanism behind the retrieval induced facilitation effect. Firstly, the older group had clear encoding difficulties (needed more learning trials to reach the 70% criterion) and were outperformed by the younger adults, even after matching performance at encoding, when tested only once at one month. While the younger and older groups which were tested repeatedly (on subparts of the initially encoded material) had equal performance. Moreover, the positive effects of retrieval were long lasting (as suggested by Chan, 2009) irrespective of age, as observed at the 1-month assessment.

The only other study of which I am aware that directly investigated repeated retrieval (of sub-parts of material) in older populations over a longer delay, is that of Baddeley, Rawlings, and Hayes (2014). The authors observed no difference in performance with a prose recall test on immediate recall but a significant overall age effect on longer delays of six weeks, although using similar integrated material to my own. They reported that the drop in performance was more clear-cut between intermediate (delays between 24 hours and 24 days) and final (6 week) delays than between the immediate and intermediate delays. Although my material and question sets closely followed the design and structure of their Crimes Test (Baddeley, Rawlings, & Hayes, 2014), I have made several changes to both the material and testing method which could account for this difference in results. Firstly, although they never used the same question twice, they did use reverse questions (i.e., "Where was the crime against the young man committed? - answer "Outside the night club." and also "What was the age and sex of the victim outside the night club?"), whereas my material never probed the same association more than once. This meant that I had to increase number of features in each story (13 as opposed to 5 in the Crimes Test). This enabled me to probe different features of the material on each of the multiple delays, thus minimising any re-learning (see Probing Scheme in Appendices). Secondly, I also selected a single delay between the post-encoding retrieval test and the following one (1 day vs. 1week). In contrast, Baddeley et al., (2014) used a range of delays and this might have generated greater variability in performance across participants, or it may be that early priming of potentially fading memories traces might be particularly effective. If the latter explanation is true, then this type of testing may prove of practical value in reducing forgetting. Several previous studies have found that the positive effects observed from repeated retrieval also relate to the length between testing intervals (Wheeler & Roediger, 1992; Chan, 2009). This offers a possible account of the discrepancy between my findings and those reported by Baddeley, Rawlings, & Hayes (2014), specifically because in my experiment I also observed a clear interaction between age and delay with faster forgetting for the older adults when memory was not refreshed by testing at one day and one week. Therefore, the results from the condition without retrieval practice provide evidence of accelerated forgetting rates in the older group, as sometimes claimed in the literature.

When replicating the first experiment using material with a disrupted narrative (Experiment 2), no statistically significant difference in performance was found between groups tested with the two kinds of material, in either the condition with retrieval practice or the condition without retrieval practice, once level of encoding was equated. I suspect that the most likely explanation is that participants tested on the material with a disrupted narrative are able to reconstruct the narrative because of a similar priming principle which applies in the case of integrated material. Even though the aim was to disrupt the narrative construction of the material as much as possible, by randomising all the sentences from the stories not just within one story, the material could still be integrated/reconstructed by participants. The reduction in forgetting during repeated retrieval is possibly also facilitated by priming that occurs at a sentence level. Given that the sentences contained multiple features, when these are probed multiple times with a different feature on each occasion, the participants might retrieve the whole of the episode within that sentence. Therefore, the only significant effect between the two types of material in Experiments 1 and 2 was that the material with a disrupted narrative was significantly more difficult to learn for most participants, in both age-groups, requiring repeated trials at encoding to reach the 70% criterion. A possible caveat with regard to priming at the sentence level is that the increased encoding time in Experiment 2 could reflect an attempt by participants to reorder the sentences to construct a coherent narrative in memory. In this case, there could still be priming of the whole narrative with the partial cueing at each repeated test. This does not undermine the conclusion that intact priming in older participants successfully removed the difference in forgetting rates between age groups.

To conclude, my experimental results are novel in several ways. Firstly, there are very few studies that have directly investigated the effect of repeated retrieval of sub-parts of material in older populations, even fewer which have done so by directly contrasting the performance of a group in conditions with and without retrieval practice. This type of design allowed for a more accurate quantification of the benefits of this testing method on long-term memory performance and also to point to the possible underlying mechanism behind this effect. Since long-term memory performance was improved in both young and old, though relearning was purposefully minimised by methodological design and older adults showed the usual age effects on learning, I infer that the repeated retrieval induces a facilitation effect does not rely on relearning mechanisms. Rather that this positive effect is better accounted for in terms of priming.

CHAPTER 3: Assessing the effect of repeated partial testing on long-term memory in a clinical sample.

As discussed in Chapter 1, much of the literature investigating the effect of repeated testing in clinical samples have typically focused on its detrimental aspects, such as the problems that arise in clinical assessment as a result of practice and repetition effects or, in the case of longitudinal clinical assessment by obscuring the underlying cognitive decline (a more detailed discussion is provided in Chapter 1, Section 1.2.3). Therefore, questions relating to the effect of repeated partial testing and its underlying mechanism in clinical populations have largely been neglected.

The fact that testing different sub parts of material (both with material with a narrative structure and one where the narrative structure was disrupted) successfully removed the difference in forgetting rates between younger and older age groups (Experiment 1 & 2) pointed to the fact that the underlying mechanism behind the partial testing effect is most likely a result of priming. As discussed in Chapter 1 (Chapter 1, Section 1.1.3), amnesic populations are also known to have learning deficits but intact priming (e.g., Camus et al., 2003; Lustig & Buckner, 2004; Bennett et al., 2006; Yano et al., 2008). Thus, if partial testing will produce the same benefits in memory performance in amnesic populations, specifically in individuals diagnosed with mild Alzheimer's disease (AD), it will provide further evidence that the effect arises as a result of priming rather than relearning.

Additionally, the methodological design from Experiment 1 & 2 allows for a proper disentangling of the effect of testing from that of actual forgetting, as it directly contrasts a condition with retrieval practice with one without retrieval practice. Therefore, this design will aid in the assessment of whether or not AD patients present with accelerated forgetting which is still an unsettled debate in the clinical literature.

The following experiment will therefore serve two purposes: 1. that of verifying whether the results from Experiment 1 can be replicated in a clinical population, namely individuals diagnosed with mild AD, thus further investigating the possibility that the underlying mechanism behind the partial testing effect is accounted for in terms of priming; and 2. assessing whether or not AD patients present with accelerated forgetting when compared to healthy age matched controls.

Experiment 3: Is forgetting in Alzheimer's Disease fast and is it affected by repeated retrieval?

3.1. Introduction.

Accelerated long-term forgetting (ALF) has been proposed as one of the main reasons for memory deficits in AD (e.g., Vallet, et al., 2016). Still, studies investigating whether AD patients present with ALF or not, have reported conflicting results (see Table 3.1). It has been suggested that these differences derive from methodological confounds (Geurts, van der Werf, & Kessels, 2015). Table 3.1 summarises the literature investigating ALF in AD and prodromal syndromes. Half of the fourteen studies I could glean from the literature found normal long-term forgetting patterns compared to those of healthy controls (HC). Several factors that could account for this discrepancy in results were identified.

Firstly, although this is not always acknowledged, a possible confounding factor is whether there are ceiling effects in the performance of HC or floor effects in the patient samples. Four out of the fourteen studies listed in Table 3.1 are marred by floor effects in the clinical sample (Kopelman, 1985 p. 634; Greene et al., 1996, p. 545; Budson et al., 2001, p. 887; Lombardi et al., 2018, p.8) while three are difficult to interpret given the ceiling effect in the control group (Greene, et al., 1996, p. 545; Degenszajn et al., 2001, p.173; Weston et al., 2018, p. 130).

Authors	**Sample**	**Material**	**Equated encoding**	**Delay**	**Recall, Recognition**	**Analysis**	**ALF**	**Floor effects**	**Ceiling effects**
Kopelman (1985)	8 AD; 14 KS and16 HC	Words (*BP test*) and Pictures (*HandP method*)	yes	10 min; 24h; 1w	recognition	percentage correct (from 10 min score)	no	yes	no
Becker, Boller, Saxton, McGonigle and Gibson (1987)	62 AD and 64 HC	Verbal passage and RCFT	no	I; 30 min	recall	difference in scores between acquisition and delayed-recall trials	no	yes	yes
Larrabee, Youngjohn, Sudilovsky and Crook (1993)	80 AD and 80 HC	Name-face associations and grocery list items	no (*but had a subset of participants who were*)	40 min (for name-face associations) and 30 min (for grocery list)	recall	difference between scores on final acquisition and delayed-recall trials	yes	no	no

Carlesimo, Sabbadini, Fadda and Caltagirone (1995)	13 AD, 8 MID, 9 Amn and 32 HC	Word list recall (*RCFT's 15 words learning task)* and Pictures (*HandP method)*	yes	I; 15 min (*RCFT)* *And* 90s; 10 min; 1 h; 24 h (HandP)	recall and recognition	comparison between recall (fifth I) and delayed trial (*RCFT's 15 words learning task)* and whole number of correct responses *(Pictures HandP method*)	yes	no	no
Hart, Robert and Others. (1987)	14 AD; 10 MAD and 14 HC	Line drawings of common objects	no	10 min; 2 h; 48 h	recognition	percentage correct	yes	no	no
Hart, Kwentus, Harkins, and Taylor. (1988).	10 AD and 13 HC	Line drawings of common objects	yes	10 min; 2 h; 48 h	recognition	percentage correct	yes at 10 min, but not at 2h and 48h	no	no

Greene et al. (1996)	33 AD and 30 HC	Prose, word list, doors and people test	no	I, 30min	recall and recognition	scaled scores	no	yes	yes
Christensen, Kopelman, Stanhope, Lorentz, and Owen (1998)	15 AD and 15 HC	Picture, forced choice word, forced choice design, a picture recall task and a stem completion task	yes (*only a subset)*	1 min; 10 min; 20 min	recall and recognition	absolute scores and percentage correct	no, for all except picture recall	no	no
Degenszajn, Caramelli, Caixeta, and Nitrini (2001)	15 AD and 15 HC	Buschke selective reminding test	no	I, 30 min; 24 h	recall	items recalled at the sixth trial of the learning phase compared to total recall at 30m and 24h	no	no	yes

Budson, Simons, Waring, Sullivan, Hussoin, and Schacter (2007)	14 AD, 19 MCI and 22 HC	Real world events	no	initial weeks after the event; three to four months later; one year	recall and recognition	proportion correct	yes, at initial weeks and 3/4-month assessment; not after one year	yes	no
Manes, Serrano, Calcagno, Cardozo, and Hodges (2008)	10 SMC; 7 SMC with objective memory impairment and 9 HC	Verbal and visual material	no	I; 30 min; 6 weeks	recall and recognition	absolute scores	yes	no	no
Walsh, Wilkins, Bettcher, Butler, Miller, Kramer, and Brown (2014)	15 MCI and 15 HC	Story learning task	yes	I (*last learning trial*); 30 min; 1week	recall	slope of the linear regression between data points	yes	no	no

Vallet, Rouleau, Benoit, Langlois, Barbeau, and Joubert (2016)	16 AD; 16 MCI and 16 HC	DMS-48	no	3 min; 1 h; 1 week	recognition	absolute scores	no	no	no
Lombardi, Perri, Fadda, Caltagirone, and Carlesimo (2018)	16 aMCI and 19 HC	120 words (*HandP*); recollection or familiarity judgements	no	10 min; 1 h; and 24 h	recognition	percentage correct	no	yes	no

Table 3.1. Summary of studies investigating ALF in AD and MCI.

AD: Alzheimer's disease; HC: Healthy controls; KS: Korsakoff's syndrome; MCI: mild cognitive impairment; aMCI: amnestic mild cognitive impairment; MID: multi-infarct demented; SMC: subjective memory complaints; BP test: Brown-Peterson test; HandP: Huppert and Piercy; Amn: amnesics; MAD: major affective disorder; RCFT: Rey complex figure test; ALF: accelerated long-term forgetting; eFAD: Presymptomatic autosomal dominant familial Alzheimer's disease; I: immediate.

Secondly, many studies failed to equate baseline performance between the clinical and the healthy group, leading to a possible incorrect assessment of the differences in the forgetting rates between the two groups. Greene et al. (1996) evaluated anterograde episodic memory in patients with AD and in HC using immediate and delayed prose recall. They reported that once initial acquisition of new information on the task was equated across groups, patients with AD did not exhibit ALF. Similarly, Kopelman (1985), using the Huppert-Piercy test, found no evidence of ALF at 24h or 7 days delay, after matching initial learning. On the contrary, Carlesimo et al. (1995) did observe ALF in AD patients at one hour and 24-hour delays on a line drawing recognition task. Recently, Weston and colleagues (2018) investigating a group of people affected by a gene mutation resulting in a form of presymptomatic autosomal dominant AD found that these people had a performance similar to HC at initial learning and 30-minute recall on a series of tests (word lists, stories, and figure recall). When assessed again after a week, people carrying the mutation had forgotten more than the non-carriers. These differences in findings cannot be attributed solely to whether initial performance was equated or not, to the type of material or testing method (recall/recognition). An additional influencing factor in investigating forgetting derives from the fact that repeated testing is inherent in the study of forgetting, but repeated testing comes with several caveats. One would be, as Weston and colleagues (2018) noted, that we cannot control for some participants rehearsing or at least recalling the material between assessments. The authors comment on the difficulties arising with repeated measures and argue for the importance of identifying new methods of assessment. They propose either to embed testing material amongst other unrelated cognitive tests, or to use recognition tests with material that would be difficult to rehearse by participants between test sessions.

Some of the previous studies have discussed the possible implications of repeated testing on patients' performance (Greene et al., 1996; Weston et al., 2018). None have, however, directly investigated the effects of such repetitions, and whether the same material or different material is used on each test session. In an attempt to address the difficulties arising with repeated testing, a number of approaches have been identified (for a review see Elliott et al., 2014). Baddeley, Allen, Atkinson and Kemp (2019) propose to use material that once learned can be used to test the same individual over longer delays, repeatedly, without testing the same information on each occasion. From the review of the 14 studies on ALF in AD, listed in Table 3.1, the issue of whether or not the same material was retested on each delay emerges as a one of the differentiating factors between studies that have reported ALF and those which have not. Six of the 14 studies that investigated ALF in AD patients, used different subsets of the initially encoded material on each testing session. These six studies documented forgetting rates in AD and aMCI similar to that of age-matched controls (Kopelman, 1985; Hart, et al., 1987; Hart, et al., 1988; Christensen et al., 1996; Vallet et al., 2016; Lombardi et al., 2018).

Lastly, as noted by Weston and colleagues (2018), repeated measures, and more importantly rehearsal do indeed raise important methodological issues. Repeated testing of the same material involves (re)learning of that material on each subsequent testing occasion. When different subsets of the initially encoded material are tested on each of the following delays, particularly if no feedback is given, then relearning should be minimised. These two types of testing procedures could lead to large differences in memory performance between individuals with learning deficits and normal groups, with healthy adults benefiting more from the relearning opportunities compared to patients. In a previous study (see Chapter 2), I suggested that memory performance benefits from repeated partial testing (testing different subparts of initially taught material) arise as a result of priming, rather than relearning. If this is to be the case, then amnesic patients should benefit to the same extent as HC as a result of repeated partial testing, thus eliminating the difference in forgetting slopes between the two groups. To surmise if repeated testing provides a new learning opportunity, individuals with learning deficits could potentially be mistaken as exhibiting ALF since they benefit from relearning to a lesser extent, compared to healthy individuals. On the other hand, if it represents priming, then patients with amnesia, such as those with AD, should also exhibit relatively preserved long-term memory performance under repeated partial testing, as the act of repetition would serve to strengthen existing representations thus also benefiting AD patients.

In a recent methodological review of ALF studies, Elliot, Isaac and Muhlert (2014) concluded that several key factors must be considered when assessing longer-term forgetting. Among their recommendations they suggest that when assessing ALF, tests should allow for repeated testing, while avoiding repeated retrieval as much as possible by using distinctive matched tests. Furthermore, standardised tests of ALF should allow for free recall and cued recall testing, or some type of testing with retrieval support. The Crimes Test (Baddeley, Rawlings & Hayes, 2014) meets both these requirements. This prose recall test is composed of four short stories, each based on an incidence of crime that contains five key features (e.g. the crime, the criminal, the location). It does not demand excessive (initial) learning time and allows for different subsamples of questions to be tested via cued recall after a range of delays. In a later study, Baddeley, Allen, Atkinson and Kemp (2019) run two experiments each comprising a repeated testing condition (testing on: immediate, 24 hours, one week and one month) and a condition involving a single test after one month. They found that both the Crimes test and a visual test showed clear evidence of forgetting in the single test condition but little evidence of forgetting in the repeated testing condition. The authors suggested that the testing of individual features (subsamples of questions) enabled participants to remember the entire episode which then acted as a further reminder. This lack of forgetting in healthy individuals could provide an ideal test of ALF by avoiding the danger of floor effects (Baddeley, Allen, Atkinson & Kemp, 2019). The current study addresses the question of whether or not ALF does characterise the memory deficits of AD patients using the procedures devised by Baddeley, Allen, Atkinson and Kemp (2019) and material which was designed by closely following The Crimes Test (Baddeley, Rawlings & Hayes, 2014).

A second question was also addressed, namely, whether the performance of AD patients is enhanced by repeated testing. Several studies have shown the advantage of repeated testing on memory performance (Carpenter, Pashler, & Vul, 2007; Pilotti, Chodorow, & Petrov, 2009; Thomas et al., 2018; Baddeley, Allen, Atkinson & Kemp, 2019). This enhancement in performance due to retesting, referred to as the testing effect, has been shown in applied situations, including educational settings (e.g., Roediger and Butler, 2011), in healthy older adults (e.g., Ferrer, Salthouse, Stewart, & Schwartz, 2004), and to some extent in individuals with memory impairments (e.g., Yan and Dick, 2006; Duff et al., 2008). While the testing effect emerges when tests probe the entire encoded material, when evaluating the effect of partial testing (probing subparts of that material) different viewpoints emerge on how this influences final memory performance. Some suggest that the benefits that arise as a result of partial testing apply only to material that can be integrated, or reconstructed by participants (e.g. prose, video as opposed to individual words, or pictures). None of the studies which directly address partial testing effects have investigated these issues in clinical samples. A more detailed review of the literature investigating partial testing in healthy samples was covered in Chapter 1 (for additional discussions also see: e.g. Chan, 2009; Chan et al., 2015; Baddeley, Allen, Atkinson & Kemp, 2019).

Some indirect evidence suggesting that repeated testing would prove beneficial to AD patients comes from reports which have shown that increasing the delays between testing when recalling information repeatedly (spaced retrieval) can improve memory performance for dementia patients and amnesiacs (e.g., Cull et al., 1996; Brush and Camp, 1998). Recalling information repeatedly has been shown to improve AD patients' performance on: object–location associations (Camp & Stevens, 1990), names of different objects (Abrahams & Camp, 1993) and prospective memory tasks (Camp et al., 1996). For example, Kinsella and colleagues (2007) investigated the benefits of spaced retrieval for improving prospective memory performance in patients with early AD compared to healthy older adults and found that the performance of most AD patients improved as a result of spaced-retrieval (combined with elaborated encoding of the task). Experiments aiming at studying retrieval practice in dementia patients have generally focused on simple cognitive tasks such as face-name associations, object-name or object-location associations, and cue-behaviour associations (see Creighton et al., 2013). The current experiment looks at a more complex task, remembering associations between multiple features within stories.

3.2. Method.

3.2.1. Participants.

3.2.1.1. Patient sample.

The patients were recruited from various geriatric institutions in Bucharest (Romania). Participants' eligibility for the AD group was restricted to patients with a diagnosis of probable AD, confirmed at 6 months follow-up, based on international diagnostic criteria (NINCDS-ADRDA: McKhann et al., 1984; DSM-IV-TR: American Psychiatric Association, 2000). Patients included in the study should have a Mini-Mental State Examination (MMSE) score between 26-18. They were assessed with a range of standard memory and global cognition tests (see Table 3.2) and with a paper version of the Temporary Memory Binding test (Della Sala et al., 2018) by the experimenter (the author of this thesis). Patients also underwent blood screening tests to exclude other potential causes of dementia, all had CT scans, and a few had MRI scans as well. Patients were excluded from the study if they had a past history of stroke, brain traumatic injury, clinical depression or alcoholism. Due to the nature of the testing material, individuals with major hearing impairments were also excluded. Written consent from all patients, or their caregivers was obtained according to the Declaration of Helsinki, as was ethical approval from the relevant ethics committees of each institution involved (Institutul National de Gerontologie si Geriatrie “Ana Aslan” București; Spitalul Universitar de Urgenta ELIAS Sectia Geriatrie Gerontologie Bucuresti; Clinica Pro-memoria Bucuresti).

Test	**AD participants' scores**		
	Range	**Mean**	**Std. Deviation**
DS (0-10)	3-8	4.6	0.9
ADL (0-10)	3-6	5.2	0.8
IADL (0-8)	2-8	5.8	2.1
CDT (0-10)	2-10	7.4	2.2
GDS (0-15)	1-14	7.7	2.6
MoCA (0-30)	10-26	18.4	4.2
TMB (0-32)	13-29	20.3	3.7

Table 3.2. AD patients' performance on the background Neuropsychological test battery.
AD: Alzheimer's disease; DS: Digit Span (Blackburn, Benton, and Shaffer, 1957); ADL: Activities of Daily Living (Katz, 1983); IADL: Instrumental Activities of Daily Living (Lawton, and Brody, 1969); CDT: Clock Drawing (Shulman, 2000); GDS: The Geriatric Depression Scale (Yesavage, et al., 1983); MoCA: Montreal Cognitive Assessment (Nasreddine et al., 2005); TMB: Temporary Memory Binding test (Della Sala, Kozlova, Stamate, & Parra, 2018).

3.2.1.2. Healthy controls.

The healthy control (HC) sample was recruited in Romania from GP surgeries and from the local communities. The GPs provided a list of older individuals who were registered with their practice whose medical files showed they were in good health. In Romania, GPs perform regular general examinations of their patients, including cognitive assessment. All the participants included in the study were healthy at the time of testing. Exclusion criteria for the HC were: the absence of psychiatric or neurological conditions, including alcohol or drug abuse or head trauma and a MMSE score higher than 28. This latter criterion was documented by GP records. Written consent from all participants was obtained.

3.2.1.3. Comparison between groups.

The initial sample included 40 patients with AD (seven men and 33 women) and 44 HC (10 men and 32 women). The HC participants were recruited to match AD patients on age, educational level and when possible gender. The AD participants ranged in age from 55 to 93 years with a mean age of 77.4 years (SD=8.4 years) while HC ranged in age from 56 to 89 years with a mean age of 75.6 years (SD=8.2 years), there was no statistically significant difference between AD and HC on age (t=.990; p=.326). The AD participants ranged from 4 to 16 years with a mean of 12.7 (SD=3.7) on level of education, and the HC ranged from 7 to 18 years with a mean of 13.5 (SD=2.8). There was no statistically significant difference between AD and HC on level of education (t=.988; p=.326).

The final sample included 33 AD patients and 42 HC. Four participants refused to take part on following testing delays (two patients and two controls); one patient had a cerebral stroke between the one week and 1-month testing delay; the performance of one patient in the condition without retrieval practice was excluded as flagged as a significant outlier and 7 patients were not included in the final analysis as they did not reach the 70% encoding criterion.

Table 3.3 details the demographics of the subgroups (AD & HC) according to experimental conditions.

GROUP		**Range**	**Mean**	**Std. Deviation**
AD Repeated Testing (N=21)	Age	55-88	75.8	8.1
	Education	4-16	11.9	4.2
	MMSE	19-26	23.4	2.4
AD Single Testing	Age	67-93	79.2	7.7

(N = 19)	Education	7-16	13.4	3.3
	MMSE	18-26	22.1	3.0
HC Repeated Testing (N = 21)	Age	56-85	73.6	7.7
	Education	8-16	13.4	2.6
	MMSE	29-30	29.5	0.5
HC Single Testing (N = 21)	Age	62-89	77.4	8.3
	Education	7-18	13.4	3.1
	MMSE	29-30	29.7	0.4

Table 3.3. Demographic variables and MMSE scores of AD and HC groups subdivided by testing condition.
AD: Alzheimer's disease; HC: Healthy controls; MMSE: Mini-Mental State Examination.

3.2.2. Design.

All testing was conducted in Romanian, all neuropsychological tests which were carried out had translated and validated Romanian versions. With regard to the Fables test, even though it was initially devised in English, it had been translated into Romanian and used in a previous experiment (Chapter 2, Experiment 1) on a large (N=120) Romanian sample of both younger and older participants.

The experiment employed a mixed design. Participants were randomly allocated to either a condition without retrieval practice or one with retrieval practice. Participants in the condition without retrieval practice were only tested at two delays: post encoding filled delay and one month. Participants in the condition with retrieval practice were tested at four delays: post encoding filled delay, one day, one week and one month.

During the encoding phase, all participants were presented with four fables read out by the experimenter at a slow and clear pace (2s pause between each sentence and 5s pause between each fable). To minimise any recency effects, each presentation phase was followed by a written one-minute filler task, involving finding as many words as possible from the letters composing the Romanian word “hippopotam” (see Baddeley, Allen, Atkinson & Kemp, 2019). Participants then took the initial post encoding filled delay cued recall test on one subset of questions (there were four subsets in total), which was self-paced. If participants scored less than 70% correct (9 out of 13 questions), the four fables were presented again (in a different order); participants took the one-minute filler task again and were then retested. The aim was to repeat this process until participants reached the 70% criterion or to a maximum of four trials.

The subsets were randomised both during the encoding phase (in the cases where more trials were needed) and across the various testing delays. In the condition without retrieval practice one of the subsets not tested at the initial test (post encoding filled delay) was randomly selected. In the condition with retrieval practice testing material changed at each delay. The encoding phase and initial test were conducted face to face while all other tests were conducted by telephone. This type of testing, telephone follow-up, has been validated by Baddeley, Rawlings and Hayes (2014) and used successfully in other studies with similar procedures (Baddeley, Allen, Atkinson & Kemp, 2019) as well as studies involving different clinical samples (Walsh et al., 2014).

3.2.3. Material.

The material was comprised of a simplified version of the Fables test previously devised for the experiments reported in Chapter 2 investigating the effects of partial repeated testing on forgetting in younger and older healthy individuals. After piloting with a small AD group, the Fables test was modified to make it more accessible for clinical use (Appendices for details). The material used in this experiment consisted of four fables loosely mimicking Aesop’s style. Each was four sentences-long and involved eight main features (i.e., characters, nationality, moral of the fable, etc.; full material in the Appendices). This generated 52 questions, which were split across four subsets. Each question in the subsets probed one sentence from each of the four fables, without ever probing the same feature twice (in the same story) within the same subset. All materials were presented in Romanian. The original Aesop’s stories are not part of the Romanian culture, not only did I select unrenowned fables, but I also enquired (some participants) at the end of the experiment if any of these were even vaguely familiar to them to ensure they were not.

3.3. Results.

3.3.1. Initial learning.

There was a significant difference between the two groups in the number of trials necessary to reach the criterion performance level set at 70% correct ($t = 7.647$, $p < .001$) with AD groups requiring more trials ($M= 2.64$, $SD= .86$) than the HC groups ($M=1.4$, SD.48). Cohen's effect size value ($d = 1.673$) suggested that the effect of group on the number of trials required to reach the 70% criterion was highly significant. Among the 42 HC, 27 required one trial and the remaining required two trials to reach criterion. Out of the 40 AD patients, two required one trial, 15 required two trials, 13 three trials and 3 four trials. Seven AD patients who did not reach the 70% criterion were excluded from the analysis. Therefore the final AD sample included in the analysis below consisted of the 33 AD patients who had reached criterion at encoding. Even after excluding the AD patients who did not reach criterion, the number of trials to reach this criterion was still not equal between AD and HC.

Mixed effects models were used to examine how groups (AD vs. HC) and testing condition (without retrieval practice, with retrieval practice) may have affected recall performance at different delays. In order to control for individual variability among participants a model assuming random intercepts and random slopes for each participant was used, and a covariance structure to account for heterogenous variances at different delays in each linear mixed-effect. Analysis of the individual forgetting curves across all AD patients and HC participants in the condition with retrieval practice revealed considerable individual variability (Figure 2.1). In terms of individual scores 3 out of 21 AD participants showed no decline between immediate and one day (1 participant improved); 6 participants showed stable or improved performance between one day and 1-week testing (5 improved); 6 participants showed no decline between one week and one month; 9 participants had a relatively stable performance across 1 day and 1-month interval. Further information on individual performance can be found in Figure 3.1.

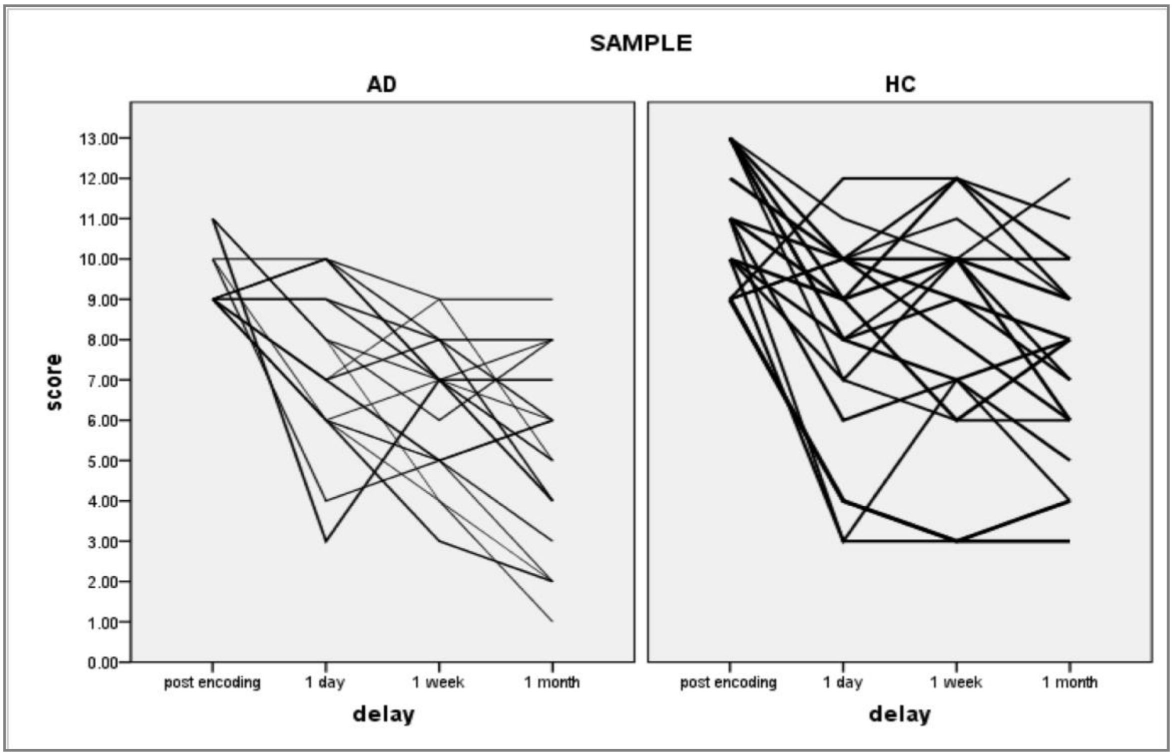


Figure 3.1. Individual recall performance on the Fables test in the immediate, 1 day, 1 week and 1-month tests in the AD and HC groups.
AD: Alzheimer's disease; HC: Healthy controls.

Random intercepts and an unstructured covariance matrix were used to account for within-subject correlations. A random effect of delay was also included in order to measure the variance in the effects of delay on scores, across participants. The significance of each fixed effect in predicting each behavioural outcome measure was assessed with $\alpha = 0.05$. A total of 248,230 data points were available for statistical analyses. Mean scores at different time intervals for each of the 4 groups are displayed in Table 3.4.

GROUP	**Delay**	**Range**	**Mean**	**Std. Deviation**
AD Repeated Testing	Immediate	7-11	9.3	1.1
	One day	3-10	7.1	2.1
	One week	3-9	6.1	1.8

	One month	1-9	4.9	2.3
AD Single Testing	Immediate	6-10	8.6	1.2
	One month	0-4	1.5	1.5
HC Repeated Testing	Immediate	9-13	10.8	1.6
	One day	3-12	8.0	2.6
	One week	3-12	8.3	2.8
	One month	3-12	7.4	2.5
HC Single Testing	Immediate	9-12	9.9	0.9
	One month	1-7	3.4	1.6

Table 3.4. Mean correct scores on the Fables test at post-encoding retrieval, 1 day, 1 week and 1-month test sessions for AD and HC groups.
AD: Alzheimer's disease; HC: Healthy controls.

3.3.2. Accelerated long-term forgetting in AD.

The first mixed effects model compared recall performance across two delay intervals only (post encoding filled delay retrieval and 1 month) between AD and HC samples, separately for each condition. The model included correct scores as the dependent variable and 2 factors: delay with two levels (post encoding filled delay retrieval and 1 month) and sample (AD and HC). Significant main effects were found in each testing condition for delay (without retrieval practice condition: $F(1, 33)= 491.851$, $p<.001$; with retrieval practice condition: $F(1, 38) = 88.360$, $p<.001$) and sample (without retrieval practice condition: $F(1, 33)= 12.441$, $p<.001$; with retrieval practice condition: $F(1, 38)= 15.345$, $p<.001$). There was no significant interaction between delay and sample in any of the experimental conditions (without retrieval practice condition: $F = (1, 33) =1.921$, $p= .175$; with retrieval practice condition: $F (1, 38) = 1.546$, $p=.221$).

Pairwise Comparisons showed that HC performed significantly better than AD at post-encoding retrieval test (MD=- 1.28 SE=.41, $p<.001$ =.004) and at 1-month test (MD= 2.32 SE= .77, $p= .005$) in the condition with retrieval practice as well as in the condition without retrieval practice (post-encoding retrieval test (MD= 0.62 SE=.26, p<.001 =.023) and at 1-month test (MD= 1.47 SE= .55, p <.001). Thus, HC participants had a significantly better performance on post-encoding retrieval test and at 1-month test compared to AD, in both conditions, however, there is no evidence of a difference between the rate of forgetting over 1-monthdelay in AD group compared to the HC in any testing condition (forgetting rates from post-encoding retrieval to 1-monthwere essentially parallel between the groups - Figure 3.2).

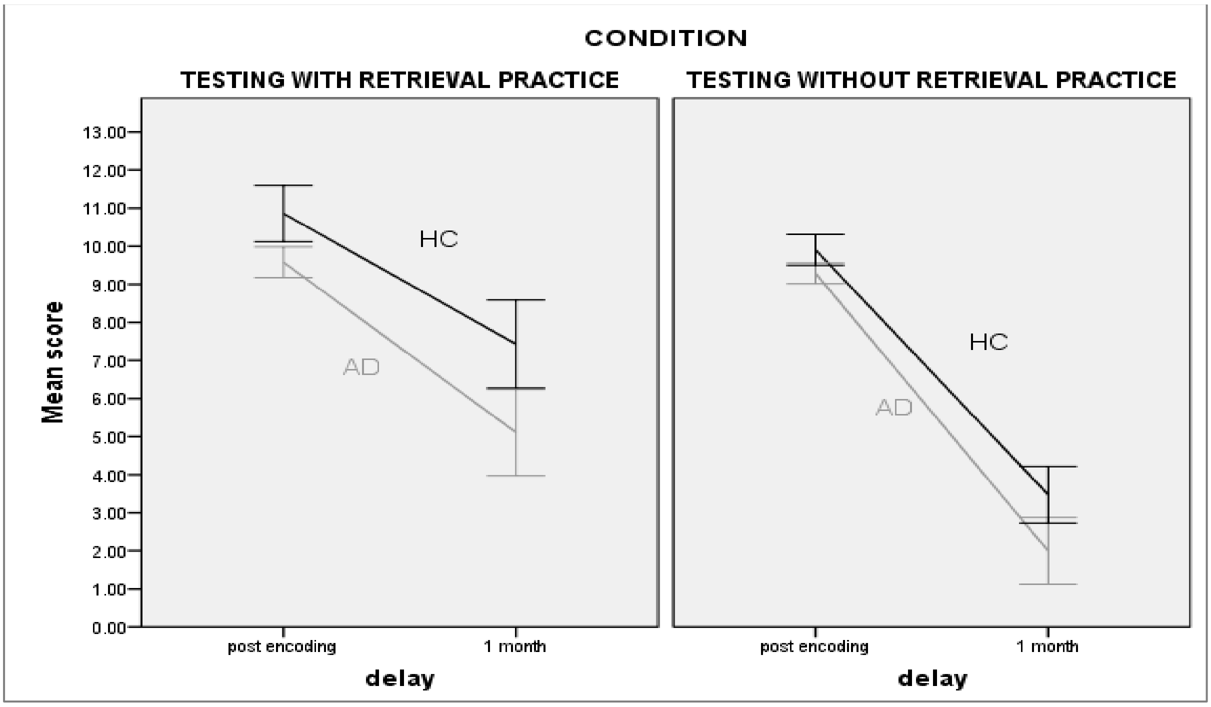


Figure 3.2. Mean recall performance on the Fables test at post-encoding retrieval and 1-month delays as a function of group (AD and HC) in both testing conditions (single testing; repeated testing).
AD: Alzheimer’s disease; HC: Healthy controls.

3.3.3. The testing effect.

A linear mixed effects model with main effects of delay, condition and sample and their interactions including the three-way interaction between all main effects as predictors was run. All three main effects, and the interaction between delay and condition, reached significance. The three-way interaction between delay, sample and condition was not significant (F (2, 71.000) =1.140, p=.326).

The second mixed effects model investigated the change in recall performance (mean correct scores) across 2 delay intervals (post-encoding retrieval, 1 month) between the 2 conditions (condition without retrieval practice vs. condition with retrieval practice). The analysis was performed separately for each group (AD, HC). Where statistically significant differences between conditions in rate of decline (i.e., a significant condition by delay interaction) were identified, model-based estimates for each delay were created.

Significant main effects were found in each sample for delay (AD: $F(1, 27) = 218.408$ $p<.001$); HC: ($F(1,40) = 185.253$ $p<.001$) and condition (AD: $F(1, 17) = 18.621$, $p<001$; HC: $F(1,40) = 35.926$, $p<.001$). There was also a significant interaction between delay and condition in each group (AD: $F(1, 27) = 10.515$ $p<.001$; HC: $F(1,40) = 35.926$, $p<.001$). AD participants in the condition with retrieval practice ($M= 5.1$, $SE= .47$) performed significantly better at 1-month($MD=3.105$, $SE=.721$, $p<.001$) compared to AD participants in the without retrieval practice condition ($M=2$, $SE=.547$) while their performance on post-encoding retrieval test was similar ($MD=.293$, $SE=.327$, $p=.416$; (AD -with retrieval practice condition: $M= 9.58$, $SE= .21$; AD -without retrieval practice condition: $M= 9.29$, $SE= .25$; $MD= .29$ $SE= .33$, $p= .416$). Three AD participants in the condition without retrieval practice performed at floor at the 1-month assessment.

HC participants in the condition with retrieval practice ($M=7.43$, $SE=.47$) performed significantly better at 1-month test ($MD=3.95$, $SE=.66$ $p<. 001$, Cohen's $d = 1.896$) than HC participants in the condition without retrieval practice ($M=3.48$, $SE=.47$), there was also a statistically significant difference in post-encoding retrieval mean scores ($MD=.95$, $SE=.40$ $p=.023$) with higher mean scores in the condition with retrieval practice ($M=10.88$, $SE= .29$) compared to HC in the condition without retrieval practice ($M=9.91$, $SE= .29$). A one-way ANCOVA was conducted with the scores from the HC group to compare the effect of condition on performance at 1-month test whilst controlling for scores on post-encoding retrieval test. Results showed that the significant effect of condition still holds ($F(14,39) = 28.092$, $p<.001$). Therefore, the HC participants in the condition with retrieval practice performed significantly better at 1-month test compared to HC participants in the condition without retrieval practice even after controlling for the differences in performance on post-encoding retrieval test (Figure 3.3).

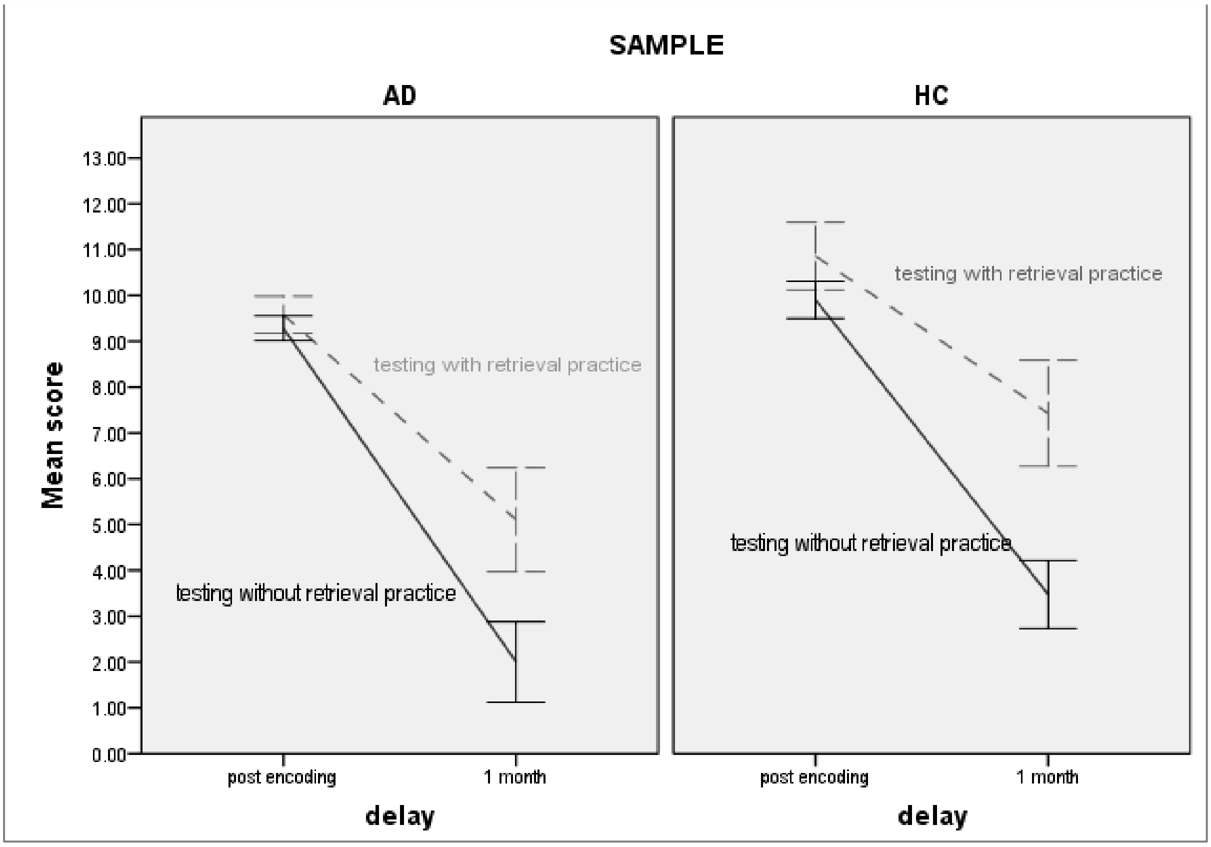


Figure 3.3. Mean recall performance on the Fables test at post-encoding retrieval and 1-month delays as a function of condition (single testing vs repeated testing) by the AD and HC groups.
AD: Alzheimer's disease; HC: Healthy controls.

3.3.4. Summary of results.

AD patients showed a significant learning deficit (requiring more trials to reach criterion) and significantly impaired recall performance on post-encoding retrieval test, as well as at 1-month test compared to HC. AD patients did not, however, show ALF between post-encoding retrieval and the 1-month test in any of the testing condition.

In both conditions both groups declined in recall performance at 1-month test compared to the post-encoding retrieval test, but the decline was significantly smaller for the groups in the condition with retrieval practice (See Figure 3.3). This suggests that repeated partial testing reduces forgetting at 1-monthdelay, producing gains in long-term retention in both AD and HC, even when retesting does not involve relearning of the tested material as different features of the initially learnt material were probed at each trial.

3.4. Discussion.

The current study had two aims: (1) to investigate whether people with AD show ALF relative to HC and (2) to investigate whether people with AD benefit from repeated testing.

3.4.1. ALF in AD.

Some authors have argued that AD memory impairment is characterised predominantly by an acquisition deficit (e.g., Kopelman, 1985; Greene et al., 1996; Grober & Kawas, 1997), whereas others have emphasised forgetting (e.g., Moss et al., 1986; Hart et al., 1988).

The AD patients in the present study did differ from HC in learning rate and showed impaired performance compared to HC at all testing delays. Patients also needed more trials to reach criterion compared to HC. Loftus (1985) has noted that differences in initial learning ability may confound analyses of forgetting rates. Other authors have also suggested that forgetting rates may be underestimated in a lower-performing group, as they have less material to forget. The present study attempted to avoid this pitfall by training all participants to a pre-set criterion (70% correct). All participants reached this criterion (after varying encoding trials), apart from seven patients who did not, and were excluded from the statistical analysis. Equating performance between patients and healthy participants can, however, present with its own limitation. Issac and Mays (1999) mention that matching procedures can in turn bias against findings that amnesiacs forget faster than controls. In order to match groups at encoding, patients invariably need longer or multiple exposures to test material compared to controls. Therefore, because the mean item-presentation-to-test delay is longer for patients, this can lead to an underestimate of the patients' forgetting rates (Issac & Mays, 1999). This design cannot exclude these possible very early consolidation differences between AD patients and controls.

The results of the present study speak against the occurrence of accelerated forgetting of verbal material in AD patients over the course of 1 month. When comparing performance from post-encoding retrieval to 1-month test, AD patients did not show ALF in either the condition with retrieval practice or the condition without retrieval practice.
When investigating ALF, a combination of recognition and free recall is recommended (Elliot, Isaac & Muhlert, 2014). The lack of a free recall measure is acknowledged as a limitation of the current experiment. A free recall measure could be easily devised for the

current test (as in the case of the Crimes test- Baddeley, Rawlings & Hayes, 2014). Free recall has, however, been proposed to be affected by disturbances of executive functions and attention that typically characterise dementia, in addition to anxiety or depression (Cerciello et al., 2017). It is also likely to reflect the level of motivation. Recognition is less affected by these variables (Cerciello et al., 2017). The present study was influenced by the Crimes Test study (Baddeley, Rawlings & Hayes, 2014) where unpublished research (Alber, 2014) showed more variability within a normal sample for free than for cued recall, presumably because cuing reduces the influence of strategy and criterion effects.

3.4.2. The testing effect.

The performance of the 33 people with AD was compared with that of the 42 age and education matched HC on the fables cued recall task. By splitting both samples into two groups based on the testing condition (condition with retrieval practice vs the condition without retrieval practice), I was able to disentangle the effect of repeated partial testing from that of forgetting, thus accurately measuring the impact of repetition on final performance. Three of the AD patients had reached floor, at 1 month, in the condition without retrieval practice. Ceiling and floor effects are considered to be a problem only if more than 15–20 % of respondents achieved either the best or worst possible score (Garin, 2014). The 3 AD patients do not represent more than 15-20% of the sample. Both AD patients and HC in the condition without retrieval practice showed significantly faster forgetting at 1-month delay compared to the condition with retrieval practice. Therefore, the condition with retrieval practice benefited both HC and AD participants.

It should be acknowledged that repeated testing is not the only factor which can affect differences in forgetting rates. Several studies have found differences based to type of assessment, e.g. free recall versus recognition (Green & Kopelman, 2002; Kopelman & Stanhope, 1997; Isaac & Mayes, 1999a), type of material, e.g. verbal versus visuo-spatial material (Lucchelli & Spinnler, 1998; Manes et al., 2005; Davidson et al., 2007) and possibly test difficulty (Freed & Corkin, 1988). Isaac and Mayes (1999a) found accelerated rates of forgetting for semantically related word lists and normal rates for free recall of lists of unrelated words in amnesics. Recognition and cued recall of both kinds of word lists appeared to decline at a normal rate. They interpret these differences in forgetting patterns as arising from impairments in long-term memory consolidation for complex associations (between 2 or more items). While the material used in the present experiment does examine complex associations (between several features), the results may only apply to material that is integrated (such as narrative) where probing one aspect of an integrated narrative might activate the entire narrative. While in the case of material with lower integration, this might not be the case. Probing subparts of material that is not integrated (such as individual words or images), may fail to prime recall of the other subparts.

Additionally, while the use of truly independent items and test forms would probably produce no benefits in performance with repeated testing, they also raise several issues. These would require more intensive initial learning time and would be more challenging to use with patients (Baddeley, Allen, Atkinson & Kemp, 2019). Several approaches to repeated testing have been adopted in previous studies. Cassel, Morris, Koutroumanidis, and Kopelman (2016) studied memory for verbal and visuo-spatial material over delays between 30 seconds and a week in TLE patients. They initially required participants to learn four separate stories and four routes, then tested retention of one story and one route per delay. Their method has the advantage of testing each item once. The drawback is a relatively heavy initial learning load, though the encoding criterion was of only six out of a possible ten correct answers. This procedure can limit potential sensitivity to scores between zero and six at each testing occasion, in some participants. A further problem is that of serial order effects during initial learning potentially favouring primacy, recency or both, which may be further complicated by test order and possible between-test interference effects (Baddeley, Allen, Atkinson & Kemp, 2019). Similarly, Jansari and colleagues (2010) tested a single patient with TLE using ten stories, testing two at each of five delays, one by recall and one by recognition. Evidence of ALF was observed that was not found when the same story was tested repeatedly. McGibbon, Firminger, and Kapur (2010) study provides important information, but requiring participants to learn ten stories would make this test impracticable with a clinical population.

Nonetheless, the fact that both AD and HC benefit from repeated partial testing to the same extent can have major practical implications. Repeated testing can thus be employed to avoid floor effects (a frequent methodological confound) in studies comparing forgetting rates between AD and HC, without compromising the validity of the comparison.

3.4.3. Conclusion.

To the best of my knowledge, this study presents the first assessment of long-term forgetting in AD patients over an interval of 1 month. It is also the first study to compare forgetting rates in AD under a condition with retrieval practice to a one without retrieval practice. By doing so I was able to uncover the importance of the number of tests and the length of test intervals when comparing forgetting rates in clinical and healthy groups over longer periods of time than have been common in previous studies.

Compared to the majority of studies on practice effects, which use within subjects' design, the current experiment employed a between subjects' design that allowed me to separate the effects of retesting from the effects of delay. Therefore, this enabled a more accurate quantification of the magnitude of this effect and showed that performance is improved under repeated testing conditions, even with partial testing (sampling different features from each fable on every test session/delay).

These results have potential practical implications in designing strategies/interventions for AD, as well as informing methodological design in clinical trials. Firstly, interventions that can be demonstrated to be efficient in aiding patients to remember important information over prolonged periods of time, are increasingly needed. Both patients and carers seek practical advice from professionals on neuropsychological interventions that will engage remaining capabilities of AD patients and are proved to promote and prolong independent functioning (Camp, 2001; Clare et al., 2002; Clare & Woods, 2004). These results offer supporting evidence that repeated testing can be used to improve AD patients long-term memory performance. Secondly, repeated testing is used in clinical assessment as well as in clinical trials and research, the evidence that repeated testing (even when only subparts of material are being tested) increases performance for both healthy and clinical patients' needs to be carefully taken into account when employing this type of design. Practice effects have been shown to result in type 1 or type 2 errors (Goldberg et al., 2015). Goldberg et al. (2015) have drawn attention to the fact that ignoring practice-effect-related gains in performance produce large sources of errors and increase the likelihood of misinterpretation of the outcomes of clinical trials.

In conclusion, this study adds to the previous literature showing that memory impairment in AD disease is primarily characterised as an encoding, or storage deficits, rather than as accelerated forgetting. It also shows that re-testing at multiple delay increases long-term memory performance compared to a single test. The beneficial effect of re-testing holds also in people with AD.

CHAPTER 4: Comparing forgetting rates between groups.

As discussed in the previous chapters, there is considerable debate in the literature regarding the comparison of forgetting rates between groups who may be performing at different levels. Failing to equate baseline performance could lead to biases in assessing the differences of forgetting rates. Group comparisons are further complicated by the effect that original learning might have on subsequent forgetting.

Starting with Ebbinghaus (1880, 1885) to the present day, several studies have shown that retention increases with the increase of intentional study trials or study time. While, whether or not the initial level of degree of learning (DOL) affects forgetting is still debated. Indeed, different experiments yielded different results. Slamecka and McElree (1983) underlined the operational distinction between retention and forgetting. While retention would be measured on a single memory test, forgetting must be measured on separate occasions with two or more assessments (Slamecka & McElree, 1983). Forgetting is hence defined as the difference among these (gradually poorer) scores and is presented graphically as the slope of a line that connects the performances at different intervals of time. Thus, they propose that any experiment analysing whether forgetting rates vary with the DOL must 'show a main effect of learning seen as an intercept difference and a main effect of interval seen as a forgetting slope' (p.384). Specifically, there must be a statistical interaction between these variables such that the slopes, or rates of forgetting, will vary depending on DOL.

One of the first exchanges on the view of whether variations in DOL subsequently affect the rate of forgetting, was between Slamecka (1985) and Loftus (1985a/b). The methods they used lead to contradictory results. For example, Slamecka (1985) concluded that the DOL has no effect on forgetting rate. Whereas, after reviewing the same literature, Loftus (1985a) reached the opposite conclusion that higher degrees of learning lead to slower rates of forgetting. Loftus (1985a) asked the question of whether the time required for memory performance to fall from one given level to a lower level was the same for different degrees of original learning. He maintained that lower degrees of learning require shorter times and lead to faster forgetting (Loftus, 1985a).

Slamecka and McElree (1983) conducted three experiments in which they examined how degree of original learning affects forgetting. They manipulated the levels of proficiency by varying the number of study trials for which participants had to learn the verbal material, either: one or three in the first experiment; two or three in the second experiment; and three or four in the third experiment. Participants were then tested at varying delay intervals, ranging from immediately after learning to 5 days. Though using different types of information (i.e., words, word pairs) and various retrieval methods (free recall, cued recall, and recognition), Slamecka and McElree (1983) obtained very consistent patterns of data, showing that degrees of original learning do not interact with delay interval. They therefore concluded that forgetting was independent of degree of original learning.

Loftus (1985) argued against defining forgetting functions as the slope between any two delay intervals. Building on Ebbinghaus's (1885) forgetting curve, Loftus (1985a/b) proposed the horizontal method of analysis to investigate degrees of learning and forgetting. That is, he assessed whether or not two functions exhibit the same rate of forgetting by testing them for parallelism in the horizontal direction. He proposed that if the two curves maintain a constant horizontal distance as a function of time, then this would reflect that both curves have the same half-life. On the contrary, if a higher DOL decreases the slope, thus slowing the half-life of forgetting, the difference between the two slopes would manifest as a continuous increase in horizontal discrepancy. Using this method, on numerous forgetting data, Loftus (1985a/b) found that higher degrees of learning result in a slower rate of forgetting. Wixted (1990) noted that the advantage of Loftus's (1985a/b) horizontal method lies in it being immune to scaling problems, which affect the usual test of vertical parallelism. He argued that this immunity derived from the fact that transforming any of the dependent measures (e.g., by squaring data) will adjust the curves in the vertical axis (direction) only and leaves differences in the horizontal axis (direction) intact.

Slamecka's and Loftus's use of different methods and definitions stemmed from their differing views on the theoretical and methodological perspectives on the relation between amount of original learning and forgetting. Yet another view on this relationship, and the consequent different method of analysis from both Slamecka and McElree (1983) and Loftus (1985a/b), was proposed by Bogartz (1990). Bogartz (1990), instead of using the previously described empirical forgetting functions, attempted to find a psychological function of forgetting. He also disagreed with the Slamecka and McElree definition of forgetting as the slope of a line. He suggested that, while a mathematical function (such as the slope of a line) may be called a forgetting function and a parameter of that function, terms such as forgetting, remembering, recalling should only be reserved for labelling psychological processes which influence observed performances (Bogartz, 1990). Despite using a different model, Bogartz (1990) findings supported those reported by Slamecka and McElree (1983), that rate of forgetting is not dependent on degrees of learning. The similarity in results between the two studies should be taken with caution, as they were not describing the same property (Wixted, 1990). Bogartz (1990) used a measure of proportional change, while Slamecka and McElree (1983) used a measure of absolute change (Wixted, 1990).

In a more recent study, Yang et al. (2016), using the same analysis as Slamecka and McElree (1983), found the opposite result, that repeated exposure to information at encoding leads to slower rates of forgetting. The comparison of time intervals in Yang et al.'s (2016) study was within participants, while in Slamecka and McElree's (1983) study it was between participants, which may have led to this difference between their results. Though within participant comparisons are extremely important, especially in informing methods for clinical studies, such as in the investigation of accelerated long-term forgetting, the repeated study/test cycles used to increase individual encoding performance raise several confounding effects. Starting with Ebbinghaus's (1885/1964) pioneering work, and in line with the work discussed in earlier chapters on the positive effects of testing on retrieval, studies have found that repetition of learning produces slower rates of forgetting.

The experiment described in this chapter is a replication of one of the classic experiments that contributed to the controversial debate of how DOL affects normal long-term forgetting which was carried out by Slamecka and McElree (1983). The data for this experiment were collected by a masters student who worked alongside me for the duration of her MSc thesis. Slamecka and McElree's goal was to find the relation, or lack of, between retention and subsequent forgetting. Their data showed that, across three experiments, study trials affected intercepts but not slopes of the forgetting functions (Slamecka & McElree, 1983).

The capacity to replicate the results of other researchers is a basic requirement for scientific integrity, however, with some exceptions, replications have not been an important part of research. The absence of replication studies is particularly problematic because empirical research is often prone to error. One variable standing in the way of replications is that there is little professional reward for carrying them out (Anderson et al., 2005). Replication studies in psychology are gaining increasing momentum, mainly motivated by the growing unease in the field with regard to possible unreliable findings. Classic studies, which are widely available in textbooks and may also be known by the general public, are particularly important to replicate (Murre & Dros, 2015). Specifically, because other representative studies, such as that of Bartlett (1932), had many unsuccessful replication attempts (e.g., Johnson, 1962; Gauld & Stephenson, 1967) prior to a successful one (Bergman & Roediger, 1999). The difference in results between replications and original studies may be due to difference in design. Such differences may be due to a lack of details, such as exact instructions, not being included in the original studies (e.g., Bartlett,1932; Ebbinghaus, 1980, 1985). Thus, replicating classic experiments, serves at least two purposes: that of verifying the reliability of the original results and uncovering with more precision how the original experiment was carried out (Murre & Dros, 2015).

Experiment 4: Replication of Slamecka and McElree's (1983) Experiment 1.

4.1. Introduction.

The current study was set up to replicate the findings of the first of three experiments reported in Slamecka and McElree's (1983) paper. As in the original study, a 2 by 4 factorial between-subjects design was used. The study material was comprised of words, chosen from the same norms categories used in Slamecka and McElree's (1983) paper. Unlike the original paper, word categories specific to a US sample were excluded and substituted with terms more appropriate for a British sample. Each participant took part in two sessions, the first was an encoding session and the second a retention test. The variables of interest were the DOL, and the retention interval (RI). DOL varied depending on the number of study trials at encoding. As in the original paper, participants were allocated either to a lower learning condition (heard the study list once), or to a higher learning condition (heard the list three times). The time interval between study and test sessions was defined as RI. While Slamecka and McElree (1983) assessed three retention intervals, immediate, 1-day and 5-days, the current study included a 4th retention interval at 10-days. The reason for including an additional interval was to investigate whether forgetting would change over a longer delay. Unlike Slamecka and McElree (1983), participants were additionally tested on The National Adult Reading Test (Nelson, 1982) which was used as a proxy for IQ. A possible limitation of this replication is that the sample used in the current study was smaller than that of the original study (70 vs 120) due to funding and time constraints in data collection. Nonetheless, I would argue that this experiment is close enough to the original to represent a replication.

4.2. Methods.

4.2.1. Participants.

Seventy-two healthy young adults were included in the study, all students recruited from The University of Edinburgh. They were offered an honorarium of £7 for volunteering their time. Inclusion criteria were English as a first language, good hearing acuity, and either being currently enrolled at the university or having graduated up to a year prior to participation. Individuals not meeting these criteria, or those with a history of clinical depression, significant psychiatric illness, dyslexia or learning difficulties were not included in the study. Of the 72 participants 24 were men and 48 women, their age range ranged between 18 and 27 (Mage=22.22, SD =1.74). Their predicted full-scale IQ (MIQ=110.9, SD =3.75) and years of education (M=17.28, SD=1.38) were as expected from the sample. Each participant took part in two sessions, a first session for acquisition and a second for a retention test.

4.2.2. Procedure.

During the acquisition session participants were allocated at random to one of two conditions. A low or a high learning condition, where the number of study trials varied, either one or three, respectively. Participants were further split into smaller groups (resulting in 9/group), each to be tested on one of four different delay intervals (immediate, 1-day, 5-days, or 10-days). All participants attended two sessions, except those allocated to the immediate testing delay who completed both the encoding and test phase in the same session. Testing took place in a quiet room with no distractions. Prior to commencing the experiment participants demographic details were recorded (age, years of formal education) and they were administered the NART (Nelson, 1982).

For each study condition, participants were instructed to listen to a list of words and try to learn its contents. One of four differently ordered lists was presented aurally to them from the electronic device. Participants were informed about the list length, that the list was blocked categorically, the number and items per category, as well as the 2-second presentation rate. They were not informed of the number of study trials or the retention test they would be given. Following Slamecka and McElree (1983), the encoding phase was followed by a 30s distractor task, writing backward subtraction by sevens from a random three-digit number. After the last encoding trial, 1 or 3 trials for the lower and higher encoding condition respectively, and the distractor task, the participants in the immediate condition were tested for retention. The remaining participants were reminded of their upcoming appointment (1, 5 or 10-days later) for the continuation of the experiment and were dismissed. Participants were not informed about the nature of the next appointment and were led to think this would not consist of an upcoming retention test on the material from their first session. The retention test (assessed with free recall) was administered exactly as in the original paper. Slamecka and McElree (1983) additionally administered participants a cued recall test after the completion of the free recall test, however, his was not done in this study.

4.2.3. Materials.

The National Adult Reading Test (Nelson, 1982) comprised a list of 50 irregular English words printed in order of increasing difficulty. Participants can only pronounce them correctly if they know/recognise them when presented in written form (Nelson, 1982). The participant has to read aloud the list of words during which the experimenter records the number of errors made. The estimated IQ is predicted from the reading error score according to the formulae by Crawford et al. (1989). The National Adult Reading Test has been found to have a high (.98) test-retest reliability (Crawford et. al., 1989).

As in the original paper, the experimental material was composed of two lists of 56 English words, drawn from the Battig and Montague’s (1969) category norms. Each list was comprised of 14 categories, with no categories in common, with 4 words from each category. The words were matched for frequency and syllable length. The presentation of the categories and the words within each category were reordered to account for primacy and recency effects. This resulted in four differently ordered lists, with 4 possible combinations for each of the 8 study conditions. The words were presented aurally, in a female voice pronouncing the words with an English accent, through a recording from a mobile device.

4.3. Results.

A post hoc power analysis indicated that having 9 subjects in each of the eight groups had 25% power for detecting a medium sized effect when employing the traditional .05 criterion of statistical significance. Two participants were removed from the final sample after conducting a Shapiro-Wilks test for normality, in each of the 8 subgroups. The test showed that the distribution of the data was significantly different from normality $p<0.01$. Thus, participants with scores of 2SD above or below the mean were removed. The subsequent Shapiro-Wilks test, after the exclusion of the two outliers, showed the data (final sample size of 70) were now normally distributed.

A 2 by 4 factorial ANOVA was conducted to investigate the effects of DOL (Higher vs Lower), RI (Immediate vs 1-Day vs 5-Days vs 10-Days), and their interaction on recall performance (the number of words correctly recalled). The alpha level for statistical decisions was set at .05. The mean recall performances at each retention interval for each learning condition are shown in Figure 4.1.

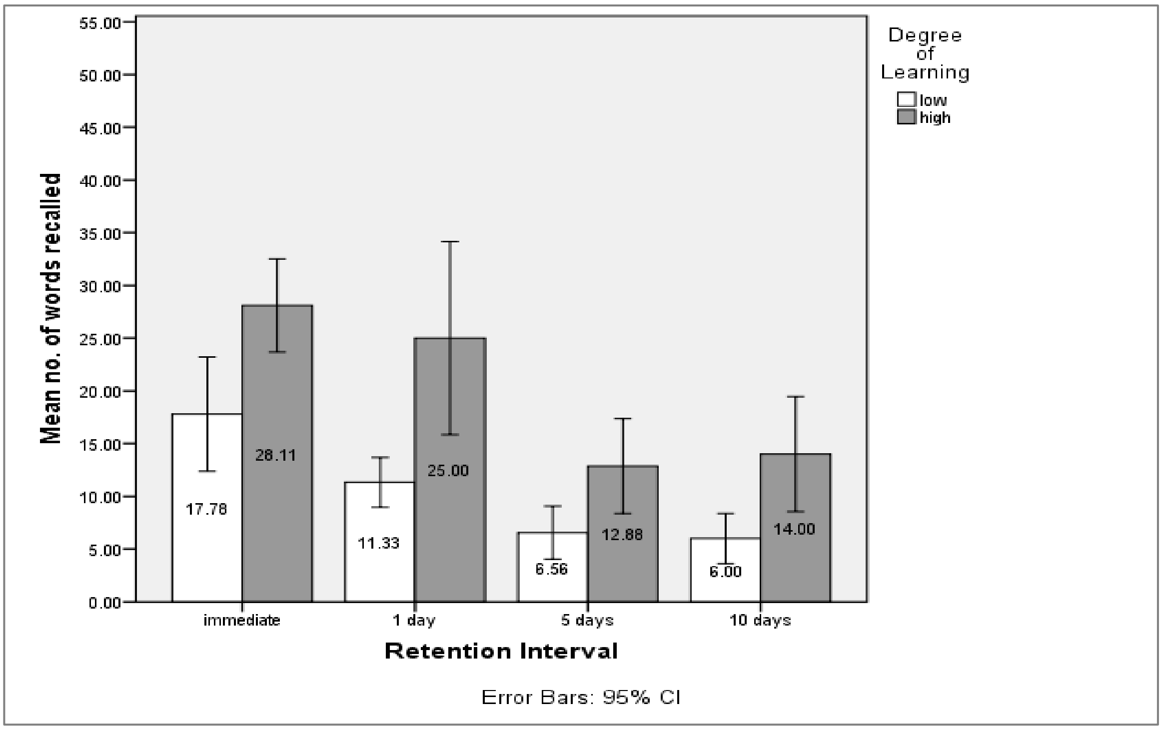


Figure 4.1. Mean recall performance at the different retention intervals subdivided according to the level of initial encoding (higher and lower degrees of learning conditions).

Results showed a significant main effect for RI (F (3,62) = 17.96, p<.001, ηp 2=.46) on recall performance. Post Hoc analyses showed that the overall mean recall performance was significantly higher at immediate testing compared to recall performance at each of the subsequent intervals (immediate-1 day: MD=4.778, SE=2.13, p=.029; immediate-5 days: MD=13.229, SE=2.17, p<.001; immediate-10 days: MD=12.944, SE=2.17, p<.001) indicating the presence of forgetting (see Table 4.1.). These results closely replicate findings of the original study by Slamecka and McElree's (1983) who also reported a highly significant effect of retention interval on recall performance. Recall performance between each retention interval was further compared, separately for the groups in the high degrees of learning condition and low degrees of learning condition. Bonferroni corrected pairwise comparisons showed that participants tested immediately after learning, irrespective of the DOL, had better performance compared to participants tested at 1 day, 5 days and 10 days but the difference at 1 day assessments was not statistically significant (immediate - 1 day comparison: low degrees of learning: MD=6.444, SE=3.015, p=.219; high degrees of learning: MD=3.111, SE=3.015, p=1; immediate - 5 days comparison: low degrees of learning: MD=11.222, SE=3.015, p=.003; high degrees of learning: MD=15.236, SE=3.107, p<.001; immediate - 10 days comparison: low degrees of learning: MD=11.778, SE=3.015, p<.001; high degrees of learning: MD=14.111, SE=3.107, p<.001). In the high degrees of learning condition recall performance decreased with each retention interval but none of the differences between two successive intervals were statistically significant.

In the high degrees of learning condition the statistical results showed that: there was a decrease in performance from immediate to 1 day, but this was not statistically significant; there was a statistically significant drop in performance between participants tested at 1 day and participants tested at 5 days (MD=12.125, SE=3.107, p<.001); and a slight, but not statistically significant, increase in recall performance at day 10 compared to day 5. The small number of participants in each group led to rather low statistical power.

Retention Interval	**N**	**Score**			
		Min.	**Max.**	**Mean**	**(SD)**
Immediate	18	9	37	22.94	(8.20)
1 day	18	7	46	18.17	(10.98)

5 days	17	1	23	9.53	(5.35)
10 days	17	2	23	9.76	(6.35)

Table 4.1. Mean recall performance (mean number of words recalled correctly) across the four retention intervals. N: number of participants.

There was also a significant main effect for DOL, ($F(1,62) = 39.17$, $p<.001$, ηp $2=.39$) on mean recall performance with participants in the higher learning condition performing significantly better compared to participants in the lower learning condition (MD=9.580, SE=1.53, $p<.001$). These results can be seen in Table 4.2. They also replicate findings of the original study by Slamecka and McElree (1983) who reported a reliable difference in performance between participants exposed to three study trials compared to those exposed to one study trial. Post Hoc analysis (pairwise comparisons with Bonferroni correction) showed that this difference in performance between participants in the high DOL condition compared to participants in the low DOL condition was statistically significant at each delay (immediate: MD=10.333, SE=3.015, $p<.001$; 1 day: MD=13.667, SE=3.015, $p<.001$; 5 days: MD=6.319, SE=3.107, $p=.046$; 10 days: MD=8.000, SE=3.107, $p=.012$).

Degrees of Learning	**N**	**Score**			
		Min.	**Max.**	**Mean**	**(SD)**
Lower	36	1	32	10.42	(6.42)
Higher	34	5	46	20.38	(10.16)

Table 4.2. Mean recall performance (mean number of words recalled correctly) by participants in the high and low degrees of learning conditions. N: number of participants; SD: Standard Deviation.

Similar to the original study, the magnitude of the effect of DOL on recall performance was not mediated by the retention interval. ANOVA showed no significant DOL by RI interaction: ($F(3, 62) = 1.09$, $p=.362$, ηp $2=.05$). Further pairwise comparisons showed that mean recall performance for the groups in the higher learning condition was greater compared to that of the groups in the lower learning condition at each retention interval (immediate: MD=10.333, SE=3.01, $p<.001$; 1 day: MD=13667, SE=3.01, $p<.001$; 5 days:

MD=6.319, SE=3.1, p=.046; 10 days: MD=8.000, SE=3.01, p=.012). Performance by higher vs. lower learner groups at each retention interval is shown in Figure 4.2. In line with the original study, these results suggest that the initial learning advantage is maintained but that there is no indication of an effect of DOL on subsequent forgetting.

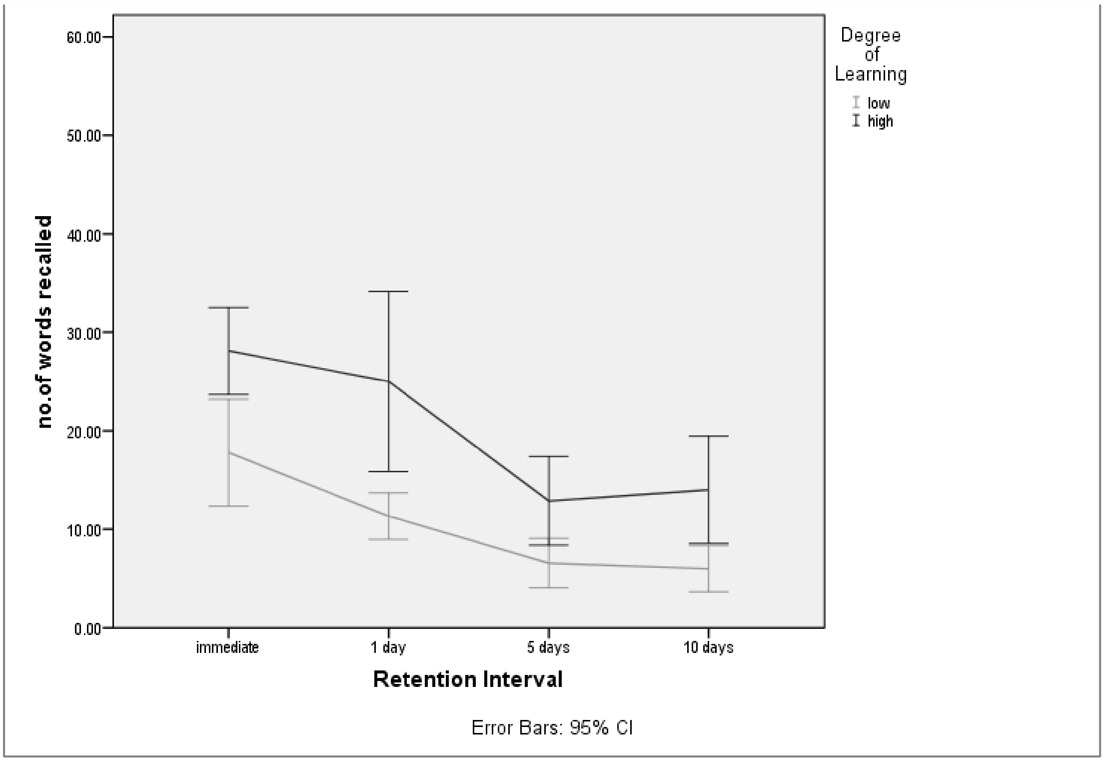


Figure 4.2. Mean recall performance by participants in the high and low degrees of learning conditions across the 4 retention intervals.

A series of Pearson correlations were conducted to observe for possible correlations between the descriptive statistics and number of words recalled. No significant correlation was observed between IQ and number of words recalled, r(68) = -0.23, p>.05 (0.059). The correlation between age and number of words recalled was also not significant, r(68) = 0.15, p>.05. Finally, there was no significant correlation observed between years of education and words recalled r(68) = 0.07, p>.05. The demographic variables were obviously rather skewed.

4.4. Discussion.

The current study closely replicates all the findings reported in the original study by Slamecka and McElree (1983). The fact that repeated learning leads to higher retention has been established since antiquity, ‘repetitio est mater studiorum’, and since proven by a large body of scientific literature (e.g., Mayer, 1983; Amlund et al., 1986; Rawson, Dunlosky, & Thiede, 2000). Whether or not repeated learning influences the rate of forgetting is an issue that has not been settled yet. In this respect, two influential viewpoints emerged from literature. Loftus’s, who proposed that when stimuli are learned multiple times forgetting rate is lower, and Slamecka’s, who proposed that multiple learning does not influence forgetting rate. This disagreement is mostly based on the different conceptualisation of ‘forgetting rate’ (Wixted, 1990). As the slope between delay intervals, in Slamecka’s case or as the time necessary for memory performance to decline from any given level to a certain lower level, in Loftus’s proposal.

The aim of the current experiment was not to settle the dispute between these two opposing viewpoints, but to reinforce the accuracy of the findings reported by Slamecka and McElree (1983). Therefore, the same methodology as their original study was employed here. The only deviation from the original was that the retention interval was extended, by carrying out an additional assessment after 10 days. The current findings are very consistent with those reported by Slamecka and McElree (1983). Participants who learned the material three times recalled more items at each retention interval compared to participants who learned the material only once, but the forgetting rate was the same. By adding a 4th assessment at 10 days we showed that the effect (in fact the lack of effect) of the DOL on forgetting rate is not moderated by the length of the retention interval. Also, similarly to Slamecka and McElree’s (1983) study, no ceiling or floor effects that may have obscured real differences in performance between groups were present. The main effects of retention interval and DOL on recall performance were large. In particular, the large effect of retention interval appeared despite the relatively low statistical power from the small size of the groups for each retention interval. The lack of an interaction between DOL and retention interval should be interpreted with caution due to the low statistical power.

A recent study by Yang et al. (2016), employing the same analyses as Slamecka and McElree (1983), provided different results. They found that learning more times slowed forgetting rate and the effect was stronger at shorter intervals. Several methodological differences may account for this divergent finding. First, Yang et al. (2016) used a within subjects’ design, which may have led to testing effects confounding the effect of learning. The lack of studies investigating how repeated learning during encoding influences forgetting is somehow surprising. Though numerous studies in educational research investigate repeated learning, most focus not on comparing different degrees of learning but rather on comparing re-learning with different methods to increase remembering, such as spaced retrieval and retesting with or without feedback.

In forgetting research, the interest in the relationship between repeated learning and forgetting is driven also by methodological considerations. Recent research has shown that equating initial performance is particularly important when comparing forgetting rates. Does repeated learning change both intercept and slope? If indeed it only increases intercept, without changing the slope of forgetting, then equating groups with different learning capacities by exposing the lower performing group to more learning trials should not affect the comparison of forgetting slopes.

CHAPTER 5: Forgetting rates and learning capacity.

The previous chapter discussed the lack of scientific reproducibility and the importance of replication studies in the field of psychology. While following the same procedures of the original study will surely increase the likelihood that scientists will successfully reproduce each other's work, in current practice most researchers conduct (replication) experiments in slightly different ways.

As previously discussed (Chapter 4), different results across studies may depend not only on the same procedures being used but also on the chosen analysis strategy, which itself is imbued with theory assumptions (e.g., Slamecka & McElree, 1983 vs. Loftus, 1985a/1985b). In the choice of both procedures and analyses, there are reasonable approaches to evaluate the data in an attempt to answer a specific research question (Carp, 2012; Wagenmakers et al., 2012; Gelman & Loken, 2014).

In the case of degrees of learning (DOL), findings from my own experiments can lead to different conclusions based on the method of testing and analysis employed. Experiments 1, 2 & 3 and the experiment which will be presented in this chapter (Experiment 5) will be used as examples. Though Experiments 1, 2 & 3 were not designed to identify whether DOL influence the rate of forgetting, such differences were assessed due to the fact that participants had to reach a pre-set learning criterion (therefore subsamples of participants were exposed to different numbers of study/test trials at encoding). Thus, in these experiments the performance of those participants requiring two learning trials (or more) was compared with that of those who took only one trial to reach criterion to see whether this influenced long-term memory performance. Based on the analysis reported in Experiments 1, 2 & 3, the results showed that the rate of forgetting is similar for participants who required one learning trial and those who required more learning trials. It is important to bear in mind that by analysing the data in that way, only the participants with a lower learning capacity (who did not reach the pre-set 70% learning criterion) were exposed to more than one learning trial. Thus only 'slower learners' were exposed to higher DOL. The next experiment

(Experiment 5), reports a (partly) contrasting result, that participants exposed to more learning trials have a slower rate of forgetting compared to those exposed to fewer trials. Thus, DOL does influence long-term memory performance, however, it does so only for faster learners (the rate of forgetting did not also slow down for slower learners in the high DOL condition). Several studies have proposed that within group variability may influence forgetting rate (MacDonald et al., 2006). Therefore, I considered that the effect of DOL on forgetting rates would be better investigated in groups that are very similar in terms of learning capacity. The next experiment was designed to further investigate the issue of DOL taking into consideration individual differences in learning.

Experiment 5: The influence of degrees of learning on long-term memory based on individual learning capacity (slower vs. faster learners)

5.1. Introduction.

Decades of research have focused on how the brain acquires and loses information (e.g., Ebbinghaus, 1885/1964; Underwood, 1964; Modigliani, 1976; Slamecka & McElree, 1983; Slamecka & Katsaiti, 1988; Squire, 1989; Rubin, & Wenzel, 1996). When assessing forgetting, specifically when comparing forgetting rates between groups with different learning abilities, a number of methodological issues need to be taken into account (Elliot, Isaac & Muhlert, 2014). A frequent methodological issue is that of matching initial learning between groups, and consequently the related confound arises due to slower learners being exposed to more learning trials (Gentile, Voelkl, Pleasant, & Monaco, 1995).

The comparison of forgetting rates between groups performing at different levels is complicated due to the lack of consensus regarding whether or not the degree of initial learning affects the rate of forgetting. Early studies have shown that learning material multiple times produces slower forgetting (Ebbinghaus 1885/1964; Loftus 1985), but as its been shown in the previous experiment (replication of Slamecka & McElree 1983, Chapter 4) other researchers find that DOL does not influence the forgetting rate (e.g., Slamecka 1985; Nilsson et al. 1989; Bogartz 1990; for review, see Wixted 1990).

A pertinent example of two studies investigating the same topic yet reaching opposing conclusions is the comparison between Slamecka and McElree's (1983) and Yang et al.'s (2016) studies. As mentioned in the previous chapter, the two studies have some methodological differences which may be responsible for the discrepancy in results. First, the comparison of delay intervals in Slamecka and McElree (1983) was between participants, while Yang et al. (2016) used a within participant comparison. Second, Slamecka and McElree (1983) used the same word pairs for recall and recognition while Yang et al. (2016) used different sets. Several studies have shown that intervening tests maintain memory performance over time, even when different subsets of the material are being tested on each delay (e.g., Butler, 2010; Chan et al., 2015; Baddeley, Allen, Atkinson & Kemp, 2019; Stamate, Baddeley, Logie & Della Sala, 2020). Therefore, in Yang et al.'s (2016) study, the effect of DOL and repeated testing are difficult to disentangle.

Though reaching different conclusions, both these studies assess the effects of DOL on performance in non-homogeneous groups. For example, in both studies a group exposed to either one or three learning trials encompassed individuals with different learning capacities (i.e., faster and slower learners). To illustrate with an example of why this may be problematic, consider giving an individual with a score of three out of a maximum of ten an additional learning opportunity. His long-term memory performance might be differently affected than that of an individual with an initial score of six. The idea that faster learners forget less than slower learners is supported by the early work by Gillette (1936) and McGeoch and Irion (1952). The latter noted that 'individual differences in learning are reflected in individual differences in retention' (p. 325). On the other hand, Underwood (1954, 1964) maintained that once the degree of initial learning is equated for faster and slower learners, there are no differences in the rate of forgetting between them. Several subsequent studies reported no differences in the rate of forgetting between slower and faster learners after equating for initial learning (e.g., Stroud & Carter, 1961; Schoer, 1962; Shuell & Keppel, 1970; Gentile et al., 1982).

More recent studies, however, have provided results in favour of the view that faster learners do have better retention over time compared to slower learners, even when initial learning is equated. Kyllonen and Tirre (1988) used an item dropout procedure to ensure equal learning for 685 participants on a paired-associates task. They reported that both item-specific learning speed and general learning speed, predicted retention. Using latent variable regression analyses, they found that cognitive ability functions such as knowledge, reasoning, and working memory did not predict retention independent of individual differences in item-specific learning speed. Their findings suggest that forgetting rate is in fact influenced by rate of acquisition.

New insight into this topic was offered by Gentile, Voelkl, Pleasant, and Monaco (1995) who proposed that while forgetting curves for faster and slower learners are the same after original learning, they differ after relearning. They had young participants (fourth and fifth graders) learn a poem to a criterion of 70-90% correct and found that both faster and slower learners recalled the material to a similar performance level after seven days. When these participants had to relearn the material, faster learners recalled significantly more than slower learners at 14 and 28 days.

A number of studies discussed the influence of individual differences in both rates of acquiring and forgetting information (e.g., Kyllonen & Tirre, 1988; MacDonald et al., 2006). These variations are further exemplified in the few studies that have reported figures of individual forgetting curves, where forgetting rates vary substantially across individuals in the same group. MacDonald et al. (2006) proposed that the analysis in aggregate change, as opposed to individual change, could account for the reason we fail to see systematic differences in forgetting rates in the existing literature. They argue that comparing mean effects is only useful for the identification of group differences, but they cannot inform us about the mechanisms and correlates that drive these differences (MacDonald et al., 2006).

Studies looking at individual performance suggest that there are large variations in participants' performance in both the intercept and slope. Wixted and Ebbesen (1997) demonstrated this by fitting power functions to individual participants' data. Their results suggested that participants did not simply differ in overall performance, rather that these differences may appear both as a result of differences in initial levels of performance and in forgetting rates. Using a paired-associates task, Unsworth, Brewer and Spillers (2011) found that individuals with higher and lower working memory capacity had similar recall levels when tested immediately, but that the individuals with lower working memory capacity showed greater forgetting when tested at longer delays. MacDonald et al. (2006) taught participants 4-digit numbers to perfection, while also controlling for possible differences in encoding strategies. They trained all the participants with the same mnemonic strategy prior to commencing the study. All participants were taught to criterion and were then retested at 30 minutes, 24 hours, 7 weeks, and 8 months later. Contrary to Underwood (1954), MacDonald et al. (2006) found significant individual differences in forgetting rates, even when equating for initial learning. In their study, faster learners forgot 58% of the items while slower learners forgot 77% of the items over the 8 month period. Zerr et al., (2018) reported that faster learners were able to retain more information over delays of 1 and 2 days compared to slower learners, despite having been exposed to less learning trials. The participants in their study were tested on a new paired-associates task (foreign-language) designed to avoid ceiling effects, by using a dropout learning procedure. They concluded that faster learners were more 'efficient learners' and that there are significant individual variations in 'learning efficiency' (the ability to both acquire and successfully retain information). Therefore, some of the more recent results suggest that there are differences in forgetting between faster and slower learners.

Zerr et al. (2018) proposed that the failure to properly identify the relationship between acquisition and retention could be attributed, at least partially, to the lack of available test sensitivity in measuring subtle individual differences in learning and retention in homogeneous populations. Because most psychometric memory tests were developed for neuropsychological purposes (i.e., detecting cognitive impairment), they tend to lack the sensitivity necessary for detecting differences in healthy younger adults (Zerr et al., 2018). This is particularly true for young adults, who often have high or near perfect performance on these tests, therefore resulting in ceiling effects which limit experimental results (Uttl, Graf, & Richter, 2002; Uttl, 2005).

The lack of consensus regarding the relation between learning capacity and retention indicates the need of equating initial performance when assessing forgetting. Therefore, we need to fully understand the possible consequences of the manipulations employed when equating initial learning in terms of how they may impact forgetting rates (Elliot, Isaac & Muhlert, 2014).

Studies directly investigating the effects of varying DOL, in the context of forgetting, have used different methodologies. Slamecka and McElree (1983) employed a between groups analysis, each group being exposed to different DOL. Because the groups in their experiment were non-homogenous, each may have included participants with different learning capacities. Therefore, each individual may have been differently impacted by relearning, making it difficult to disentangle the effects of DOL and retention capacity within a group. Other experiments have classified slower and faster learners by counting the number of learning trials required to reach a certain learning criterion, and then compared the forgetting curves between faster learners and slower learners. Such methodology cannot properly quantify the effect of DOL on subsequent retention within a group. A more robust method for doing so would be to compare participants with the same initial learning capacity exposed to different numbers of learning trials.

Therefore, the purpose of the current study was to examine this effect, using a different methodology. Experiment 5 proposes that the reason some studies found that DOL has no influence on performance at a later delay is mainly because they allocate participants with varying initial scores to different DOL. In the current experiment, in order to minimise individual differences in acquisition, and their possible influence on long-term retention within groups, participants with the same initial performance were exposed to both high and low learning conditions. I then compared how the learning condition influenced forgetting rates (without the confounding influence of individual variance in acquisition). Lastly, Experiment 5 uses material that had been previously tested in a large sample (Experiment 2, Chapter 2) from which it seemed unlikely that ceiling effects would be observed, even with multiple learning trials. This previous experience with the material also allowed for an estimate, based on the previous samples, of the range of scores that should be used as a cut-off for both faster and slower learners (e.g., choice of score for slower and faster learners was based on what the 'typical' higher and lowers scores were, using this material in similar samples).

5.2. Methods.

5.2.1. Design.

The experiment employed a between-subjects design, and included three independent variables, all varying between participants. Factors included learning category (slower vs. faster learners), DOL (higher vs. lower), and retention interval (post-encoding retrieval vs. 1-monthdelay).

5.2.2. Material.

The material consisted of the scrambled sentences from the unfamiliar fables by Aesop, the same material that was used in Experiment 2 described in Chapter 2 (see full material in the Appendices). The material can be tested using one of the four subsets. Each question in the subsets probes a single sentence, without ever probing the same sentence within the same subset.

5.2.3. Procedure.

During the encoding phase, all participants were presented with the 16 sentences, these were read out by the experimenter (author of this thesis) at a slow and clear pace (2s pause between each sentence). To minimise any recency effects, each presentation phase was followed by a 1 minute filler task that involved creating as many words as possible from the Romanian word “hippopotam” (see Baddeley, Allen, Atkinson & Kemp, 2019). Participants then took the initial post-encoding cued recall test on one of the four subsets, which was self-paced.

Participants were then classed, based on their initial test score to one or two types of groups. Those participants who scored 6 or 7 (out of 13 questions) were classed as slower learners and those who scored 9 and 10 were classed as faster learners. Participants from both subgroups (classes: slower learners and faster learners) were then randomly allocated to one of two conditions. A condition with a lower DOL and one with a higher DOL, one and two trials at encoding respectively. An encoding trial was composed of the presentation phase, followed by the 1 minute filler task and the cued recall test. Thus, participants in the higher DOL condition, were exposed to two such trials.

This procedure resulted in 4 subgroups, slower learners in the low DOL condition, faster learners in the low DOL condition, slower learners in the high DOL condition and faster learners in the high DOL condition. Participants from all the subgroups were tested at two delays: post-encoding and after one month. On each delay, the test involved a different subset of questions. The encoding phase and initial test were conducted face to face while the final assessment was conducted by telephone. This type of testing has been validated in previous studies (e.g., Baddeley, Rawlings & Hayes, 2014; Baddeley, Allen, Atkinson & Kemp, 2019; Stamate, Baddeley, Logie & Della Sala, 2020). All testing was conducted in Romanian (first language of this dissertation’s author and all volunteers).

The choice of scores, 6 or 7 correct answers for slower and 9 or 10 correct answers (out of a total of 13) for faster learners was based on previous experience with using this material in a similar sample (Experiment 2 in Chapter 3). The maximum score of 10, for the faster learners was chosen to avoid ceiling effects in the high DOL group.

5.2.4. Participants.

A total of 145 young participants were recruited mainly from Carol Davila University of Medicine and Pharmacy of Bucharest and Politehnica University of Bucharest, Romania, and a few from among friends and acquaintances. Out of the total number of participants recruited, 52 did not meet the scores for inclusion in either of the two groups (slower or faster learners). The 52 excluded participants had scores of 8, scores of below 6, or scores above 10. Out of the 93 who were included, nine participants dropped out, and did not answer the phone for the 1-month assessment. The final sample included: 20 faster learners and 24 slower learners in the lower DOL condition and 20 faster learners and 20 slower learners in the high DOL condition. Participants' written consent and demographic information concerning education, gender and age were obtained before starting the experiment.

5.3. Results.

Mean recall scores on all testing sessions for the slower learners and faster learners for each encoding condition (lower and higher DOL) are displayed in Figure 5.1 and Table 5.1. The performance for each group is displayed in Figure 5.4a, b, c, d (see Appendices).

Group	Delay	N	Min. Score	Max. Score	Mean	Std. Deviation
Slower Learners Low DOL	Encoding trial 1	24	6	7	6.38	0.49
	Encoding trial 2	24	N/A	N/A	N/A	N/A
	One month	24	1	8	4.29	1.9
Slower Learners High DOL	Encoding trial 1	20	6	7	6.6	0.5
	Encoding trial 2	20	6	12	9.2	1.64
	One month	20	1	9	5.15	2.23
Faster Learners Low DOL	Encoding trial 1	20	9	10	9.6	0.5
	Encoding trial 2	20	N/A	N/A	N/A	N/A
	One month	20	1	7	3.95	1.82
Faster Learners High DOL	Encoding trial 1	20	9	10	9.4	0.5
	Encoding trial 2	20	8	13	10.25	1.71
	One month	20	6	11	7.8	1.76

Table 5.1. Mean recall scores for slower and faster learners in each encoding condition (lower and higher DOL) on all testing delays. N/A: not applicable, no second encoding trial for low DOL groups. DOL: degrees of learning.

A three-way Mixed ANOVA with α =0.05, was conducted to examine the effects of DOL (low DOL, high DOL), learning capacity (slower learners, faster learners) (between factors) and any potential interactions on recall performance (mean recall score) at post-encoding retrieval and 1-month assessments (delay -within factor). For the groups in the high DOL condition the mean score obtained after the second encoding was used in the analysis. Mean recall performance was significantly lower at 1-month compared to post-encoding retrieval assessment (significant effect of delay [$F(1, 80) = 281.595$, $p<.001$, $\eta p2= .779$]); there was a significant main effect of learning capacity [$F(1, 80) = 33.389$, $p<.001$, $\eta p2=. 294$] with overall higher recall performance by the fast learners and a significant delay by learning capacity interaction [$F(1, 80) = 5.376$, $p=.023$, $\eta p2= .063$] suggesting that the change in scores over time is different depending on learning capacity.

The main effect of DOL was significant [$F(1, 80) = 51.592$, $p<.001$, $\eta p2= .392$] but the delay by DOL interaction was not significant [$F(1, 80) = 2.114$, $p=.150$, $\eta p2=. 026$] suggesting that the effect of learning occurred at both assessments (post-encoding retrieval and 1 month). The three-way interaction delay by DOL by learning capacity was statistically significant [$F(1, 80) = 37.105$, $p<.001$, $\eta p2= .317$].

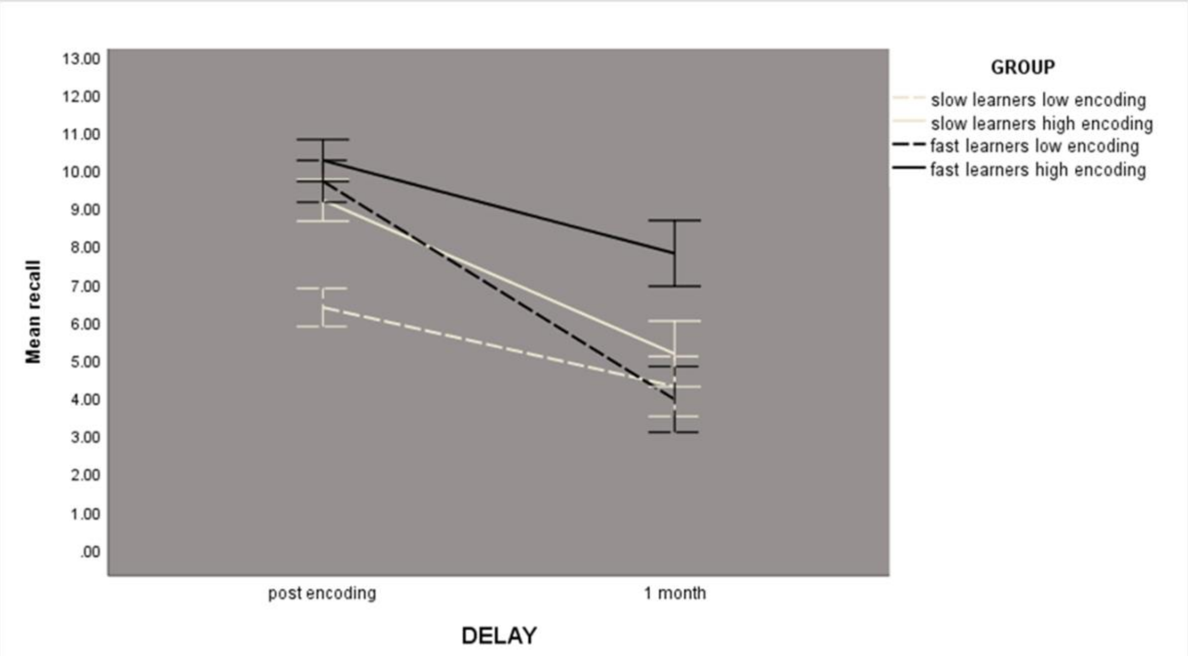


Figure 5.1. Mean recall performance at post-encoding retrieval and 1-month assessment as a function of learning capacity and DOL.

To fully characterise the three-way interaction, we further investigated the effect of DOL (by performing subsequent two-way delay by DOL mixed ANOVAs split across levels of learning capacity: slower learners, faster learners) and the effect of learning capacity (by performing two-way delay by learning capacity mixed ANOVAs split across levels of DOL (high DOL, low DOL). All two ways ANOVAS were followed by post hoc analyses Bonferroni corrected pairwise comparisons at each delay (post-encoding retrieval, 1 month). For the groups in the high DOL condition the mean scores obtained after the second encoding were used as the post-encoding retrieval mean score.

5.3.1. The effect of DOL.

Results of the delay by DOL mixed ANOVA for the slow learners' groups showed that performance at 1-month assessment declined significantly compared to post-encoding retrieval assessment (significant effect of delay [$F(1,42) = 100.922$, $p<.001$, $\eta p2= .706$] in both slower learners' groups: MD= -2.083, $p<.001$ for the slower learners in lower DOL group and MD= -4.050, $p<.001$ for the slower learners in higher DOL group). The slower learners' group exposed to the second encoding trial (higher DOL condition) had a significantly better recall performance at post-encoding retrieval assessment compared to the slower learners' group exposed to a single encoding manipulation (significant main effects of DOL [$F(1,42) = 20.322$, $p<.001$, $\eta p2= .326$; (MD=2.825, $p<.001$)]) but both groups performance was similar at 1 month (MD=858, $p=.180$), explaining the significant delay by encoding interaction [$F(1, 42)= 10.377$, $p=.002$, $\eta p2= .198$]. Therefore, the effect of the second encoding trial was lost by the 1-month assessment for the slow learners.

The delay by DOL mixed ANOVA for the high learner groups showed that, similar to the slower learner groups, performance at 1-month assessment declined significantly compared to the post-encoding assessment (significant effect of delay [$F(1,38) = 192.524$, $p<.001$, $\eta p2= .835$] in both conditions: MD= -4.050, $p<0.01$ for the faster learners group in the low DOL condition, MD=-2.450, $p< .001$ for the faster learners group in the high DOL condition). The effect of the second encoding manipulation was significant [$F(1,38) = 32.620$, $p<.001$, $\eta p2= .462$], with higher overall performance by the group in the high DOL condition. The delay by encoding interaction was significant [$F(1, 38) = 30.048$, $p< .001$, $\eta p2= .442$] and was explained by the fact that the faster learners group exposed to the second encoding had similar recall performance at the second post-encoding retrieval assessment as the faster learners group exposed to a single encoding (MD=.550, $p= .195$) but significantly higher mean recall scores at 1-month assessment (MD=3.850, $p<.001$).

Results of the two mixed ANOVAS therefore show that a higher DOL (the second encoding manipulation) decreased forgetting at 1 month for the faster learners but not for the slower learners.

Pearson correlations were performed separately for participants in the low DOL condition and participants in the high DOL condition. For the participants in the high DOL condition there were strong positive correlations between the score at the first post-encoding retrieval assessment and the score at 1 month ($r= .598$, $p < .001$ – significant at the 0.01 level, 2-tailed), between the score at the first post-encoding retrieval assessment and the score at the second post-encoding retrieval assessment ($r= .402$, $p =.001$ – significant at the 0.05 level, 2-tailed), and between the score at the second post-encoding retrieval assessment and the score at 1 month ($r= .518$, $p =.001$ – significant at the 0.01 level, 2-tailed). For the participants in the low DOL condition there were no significant correlations between the score at first post-encoding retrieval assessment and the score at 1 month.

5.3.2. The effect of learning capacity.

Results of the mixed ANOVA for the groups in the high DOL condition with mean recall scores on three delays (first post-encoding retrieval assessment, second post-encoding retrieval assessment and 1 month) as within-subjects factor and learning capacity (slower learners and faster learners) as between-subjects factors, showed a significant main effect of delay [$F (2, 76) = 62.631$, $p< .001$, $\eta p2= .662$], a significant main effect of learning capacity [$F (2, 76) = 5.572$, $p= .006$, $\eta p2= .498$] and a significant delay by learning capacity interaction [$F (1,38) = 37.697$, $p<.001$, $\eta p2= .128$]. Bonferroni corrected pairwise comparisons showed that the faster learners group performed significantly better compared to the slower learners group on first post-encoding retrieval assessment (MD=2.800, $p<.001$). Performance improved significantly from the first post-encoding retrieval assessment to the second encoding for the slower learner’s group (MD= 2.600, $p< .001$) but not for the faster learner’s group (MD= .850, $p= .060$) therefore after the second encoding slower learners and faster learners had similar performance (MD=1.050, $p= .055$). Mean recall performance declined significantly at 1-month assessment for both the slower learners’ group (MD= -4.050, $p< .001$) and the faster learners’ group (MD= -2.450, $p< .001$) but the decline was steeper in the slower learners’ group which had a poorer performance at 1 month compared to the faster learners group (MD= -2.650, $p<.001$).

Therefore, in the higher DOL condition learning capacity did influence performance at 1-month assessment, with faster learners forgetting less than slower learners (Figure 5.2).

Comparing slower and faster learners in the low DOL condition showed no significant difference in mean recall at 1 month. There was a significant overall decline from the post-encoding retrieval assessment to the 1 month assessment (significant effect of delay [$F(1, 42) = 193.117$, $p<.001$, $\eta p2= .811$] in both groups (slower learners: MD=-2.083, $p<.001$; faster learners: MD=-5.750, $p<.001$)).The faster learners group had significantly better mean recall performance compared to the slower learners group on post-encoding retrieval assessment (MD= 3.325, $p< .001$) but the difference was lost at the 1 month assessment (MD= .324, $p= .553$) where faster learners and slower learners groups had similar performances. Therefore, the significant delay by learning capacity interaction [$F(1,42) = 42.313$, $p<.001$, $\eta p2= .477$] is explained by the significant difference at post-encoding retrieval assessment.

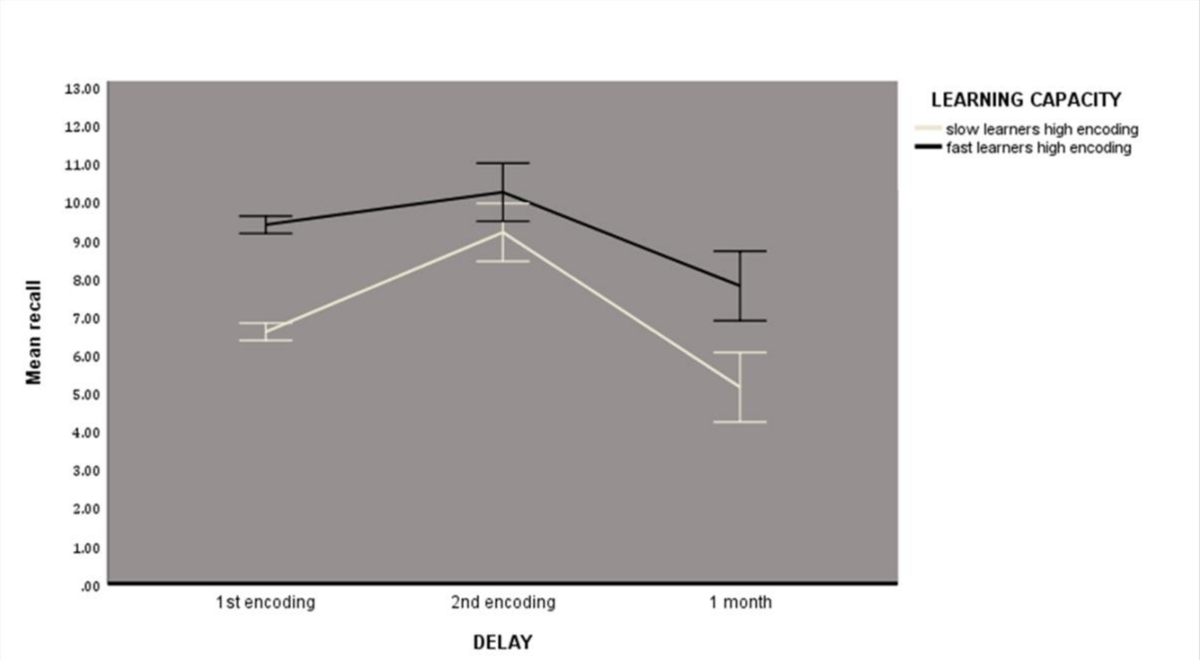


Figure 5.2. Mean recall performance after the first encoding trial, after the second encoding trial and after the 1-month assessment, by the faster learners group and slower learners group in the high DOL condition.

5.3.3. Item (feature) analysis.

Cross tabulations between first post-encoding retrieval and second post-encoding retrieval assessments and between the second post-encoding retrieval and 1-month assessments showed that the qualitative difference between slower learners and faster learners in the enhanced encoding condition is that all the feature associations that were gained after the second encoding by the slower learners group were not consolidated as these were no longer recalled at 1-month assessment (Figure 5.3).

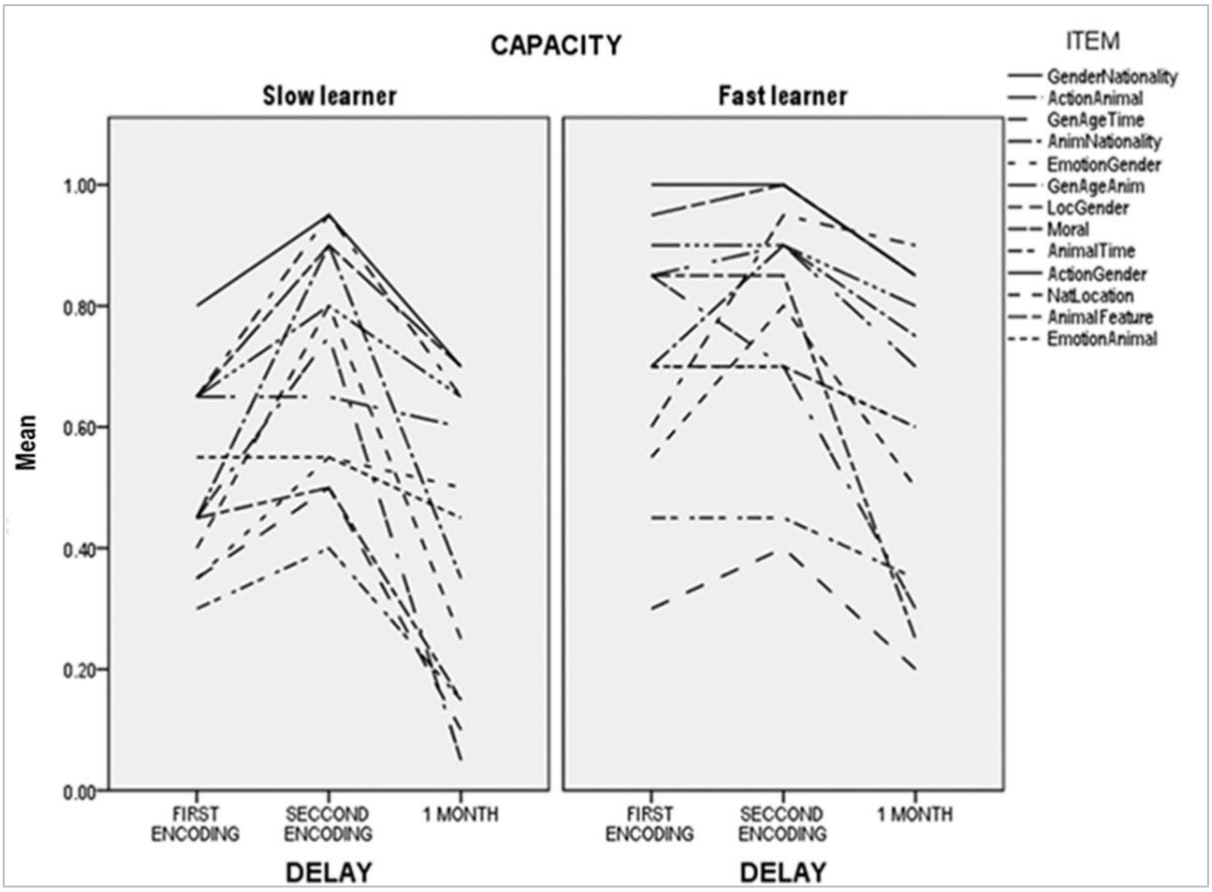


Figure 5.3. Mean number of accurate recalls for each feature-association at first post-encoding retrieval assessment, second post-encoding retrieval assessment and 1-month assessment by slower learners and faster learners in the high DOL condition.

5.4. Discussion.

Two different viewpoints emerged from the research assessing the relation between acquisition and retention capacity in the context of forgetting. One viewpoint is that faster learners not only acquire but also retain more information (slower forgetting, e.g. Zerr et al., 2018). Another viewpoint, first proposed by Underwood (1954), is that the better retention of faster learners may in fact reflect differential degrees of initial learning. Once degree of initial learning is equated for faster and slower learners there are no differences in the rate of forgetting between them. Slamecka and McElree (1983) further reported that varying degrees of initial learning has no influence on forgetting rates. While a later study by Gentile et al. (1995) showed that while forgetting curves for faster and slower learners are the same after original learning (i.e., learning to criterion 75-90% correct), they are different after relearning, with faster learners outperforming slower learners.

The results of the present study partly replicate some of these previous findings. When comparing recall performance between participants exposed to one versus two learning trials forgetting slopes were similar for the groups. A higher DOL (two learning trials) increased the intercept but did not affect the forgetting slope. Investigating the effect of DOL separately in the slower learners and faster learners groups revealed that the effect of relearning was, however, different in slower learners compared to faster learners. Slower learners in the present study had similar performance at one month, irrespective of the number of trials at encoding (DOL). While a second encoding trial did improve short-term recall performance for the slower learners in the high DOL condition, it did not decrease the slope of forgetting compared to slower learners in the low DOL condition. This pattern of results is in accordance with Underwood (1954) and several other studies (e.g., Slamecka and McElree, 1983; Meeter, Murre, & Janssen, 2005).

When analysing the performance within the faster learners groups a different pattern is observed. In the case of the faster learners, the second learning trial did not lead to significant gains in recall performance (even if there was no ceiling effect) as compared to the first post-encoding retrieval trial, but it did lead to improved performance at the 1-month assessment compared to the faster learners exposed to a single trial, as well as compared to slower learners in the high DOL condition. When comparing slower learners and faster learners exposed to a single encoding trial, though faster learners recalled significantly more feature associations at the post-encoding retrieval assessment, the advantage was lost at 1 month when faster and slower learners recalled approximately the same amount of feature associations.

It might be that the difference in results between studies showing no influence of DOL and those that do lies in the large variability of individual performance within groups (both those exposed to low and high DOL). MacDonald et al. (2006) suggest that the failure to detect differences in forgetting rates may be a consequence of various experimental characteristics, including among others the focus on between groups differences as opposed to within participant change. While the current experiment also does not specifically focus on individual change, in its case, there is less variability within groups, because participants were selected a-priori, based on initial scores, and were only then allocated to both high and low DOL.

With regard to an explanation as to why slower and faster learners are affected differently, while a firm conclusion cannot be drawn, after running an item (feature-association) analysis the pattern of results would suggest that the main difference between them is driven by how they consolidate the category of items which are encountered on a second trial. Specifically, while faster learners retain the items which are encountered on the second trial up to the 1-month assessment, slower learners forget these items, but not the ones gained on the first trial. Gentile and colleagues (1995) propose that, a plausible prediction is that faster learners will show a memory advantage when afforded the opportunity to reorganise previously encountered material, as they are better at finding strategies to organise material. While the present study did not investigate the issue of differences in individual strategies used, when debriefing participants most high scorers reported having identified some mode of organising the material after the first study/test trial. As we only see a difference in performance between encoding trials for slower learners, and not faster learners (no statistically significant difference between 1st and 2nd trial) another possibility is that, slower learners allocated their resources to learning items not encoded during the first encoding trial. While faster learners, having less new items to learn, were able to allocate more resources to consolidate already encoded items.

A limitation of the current study is that of a possible regression to the mean in the lower DOL group. What is clear, is that with this material, slower learners in the high DOL condition only benefited in the short-term (higher performance on the second trial) but not in terms of long-term memory performance. Additionally, in the high DOL condition, though slower learners reached a similar level of performance after the second encoding trial, performance at 1-month was still higher in the faster learning group. This is a useful finding as Experiment 5 was able to show that scaling problems can be avoided, by bringing all participants to the same initial score, while at the same time not changing the slower learner’s long-term memory performance.

CHAPTER 6: Forgetting in different memory systems. Assessing the effect of repetition on long-term prospective memory performance.

Much of the current thesis has focused on long-term forgetting in the context of retrospective memory (RM). Yet an important classification of long-term memory is based on the temporal direction of the memories. While RM deals with remembering of past information (e.g., learned words, events), prospective memory (PM) deals with the content of what is to be remembered or remembering to perform an intended action in the future. Further, PM can be distinguished between event based and time-based, and is often triggered by a cue. A common assumption in the literature is that PM tasks are similar to cued recall tasks in RM (McDaniel & Einstein, 1993). The similarity between PM and RM tasks has at times led prominent scholars to wonder if PM is distinct form of episodic memory (Crowder, 1996; Roediger; 1996).

An important distinction between these two types of memory, specifically with regard to their evaluation, must be taken into account. In retrospective remembering in real life, retrieval is not always prompted by a request to remember (Ebbinghaus, 1964); however, in laboratory investigations of RM, there is a request for the participant to attempt to recollect prior information. Tulving (1983) formulated that during RM tasks the participant is specifically directed to be in retrieval mode. In contrast, in PM tasks, in both laboratory and everyday settings, there is rarely a request for a memory search, rather recollection of the intended action (at the appropriate time) somehow occurs without some agent stimulating retrieval (see Einstein & McDaniel, 1996). Thus, in a PM task, attention somehow needs to be switched from the ongoing task to thinking about the intended action and performing it (McDaniel & Einstein, 2000). This following study investigated whether the recollection of an intended action at the appropriate time would be facilitated if participants repeated back the PM task multiple times at encoding (i.e., learned the PM task more). Additionally, the experiment was set up to investigate whether cued recall performance correlates with PM performance, when assessing both in the long-term (over a 1-month delay). It has been proposed that the delay that precedes performance on a PM task imposes a demand on long-term memory (i.e., storage and retrieval of intentions from a permanent memory). Thus, Experiment 6 explores the hypothesis that a positive relationship between participants long-term memory performance and PM performance may be found.

Experiment 6: Enhancing episodic memory performance over the long-term through repetition of the PM response

6.1. Introduction.

PM is responsible for many parts of everyday cognition, failures in PM can be as debilitating as those in RM. An important assumption in the literature is that episodic PM tasks are similar to cued recall tasks in RM (McDaniel & Einstein, 1993) as both tasks require forming an association between two pieces of information. For example, in the case of cued recall between two words, or between a word and a cue. In the case of event-based PM the association would be between the intention and the action to be performed, requiring to later reinforce this association at retrieval.

The similarity between PM and RM tasks has at times led prominent scholars to wonder if PM ought to be kept distinct from RM episodic memory (Crowder, 1996; Roediger; 1996). Crowder and Roediger questioned the notion that PM is a distinct form of episodic memory. While it is important to note they did not reject the idea that memory has a prospective function, they did question whether the function involved distinct processes and structures (e.g., different cognitive system) from those required in episodic RM (Graf, 2001). The methods used for assessing performance on PM tasks limit research from reaching a conclusion regarding this distinction, as well as regarding the ecological validity of laboratory findings, especially in older populations (Rendell & Craik, 2000; Verhaegen, Martin, & Sedek, 2012). Graf (2001) proposes two attributes that limit the conclusive assessment in distinguishing between PM and RM: the fact that existing methods yield binary data, and that PM performance is not ‘process pure’ (p. 539). PM data tend to be less precise and more variable compared to continuous data obtained from standard episodic RM tests, due to the fact that performance is usually scored either as a success or a failure (Graf, 2001). With regard to the second attribute, of PM tasks not being ‘process pure’, this can be explained using the previous example. One can fail on the task either due to a failure in recollecting the intention (e.g., stop when passing the supermarket) or, by not remembering the action to be performed (e.g., buying groceries). Only the first of these two components is truly prospective (Dobbs & Rule, 1987; Einstein & McDaniel, 1996) while the latter can be conceived as RM, similarly to failing to recall a list of words when cued. Remembering what the activity to be performed (the content) is considered the retrospective component of a PM situation, while first remembering to perform the activity at the appropriate occasion is what specifically defines a performance as 'prospective' (Brandimonte, 1991).

Brandimonte (1991) proposed that the PM process is composed of six steps: forming an intention; remembering 'what' that intention is (content); remembering 'when' to perform it (content); remembering to perform the action; performing the action at the right time, place, and in the way prescribed (compliance); remembering having performed it, so as not to repeat (cancellation stage). One critical aspect that differentiates PM from RM tasks is that participants self-initiate retrieval of an intended action at a specific moment, while a successful RM tasks requires externally prompted retrieval of past information (Ellis & Kvavilashvili, 2000; Kliegel, McDaniel, & Einstein, 2007). How this self-initiated retrieval of the PM response/action occurs is one of the most debated theoretical issues in the context of PM. Three theoretical perspectives have emerged from the literature.

The preparatory attentional and memory theory (PAM), which proposes that in order to successfully retrieve a PM intention we need to monitor the environment for the appropriate target cue (Smith, 2003; Smith & Bayen, 2004). Engaging in constant monitoring processes would require a certain amount of attention and working memory resources, which would consequently induce costs on the ongoing tasks. The multiprocess theory assumes that the retrieval of a PM intention may employ, across the different stages of the PM task, either monitoring or spontaneous retrieval, or an interplay between both types of processes (McDaniel & Einstein, 2000; Einstein et al., 2005). The 'choice' of process will depend on various factors such as: the target cue (event) focality, the duration of the retention interval, the importance of the task, the nature of the PM task, the ongoing task/activity performed during the retention period, as well as personal characteristics (McDaniel & Einstein, 2000; Marsh et al. 2003; Scullin et al., 2013). For example, when the PM task is important, or frequent, people are more likely to use monitoring strategies. When a cue is focal to the ongoing task or when the delay between encoding and retrieval of the PM response is long, people are more likely to rely on spontaneous retrieval. In everyday life particularly, one can employ both spontaneous and monitoring strategies to retrieve a PM intention. For example, passing by a drug store can spontaneously retrieve the memory that you have to buy something this would trigger a search in the memory for what particular thing you have to buy (Meier, Zimmermann & Perrig, 2006). Whereas the theoretical background of the different stages of a PM tasks are provided by these theories, the practical issue of what and how to improve self-initiated retrieval for the PM intention remains to be established.

Several characteristics of the PM cue such as distinctiveness (McDaniel & Einstein, 1993; Brandimonte & Passolunghi, 1994), familiarity (Watkins & Watkins, 1975, Anderson, 1985 cited in Einstein & McDaniel, 1990) saliency, specificity, emotional value, have been found to have a positive effect on PM performance (Brandimonte & Passolunghi, 1994; Einstein et al., 1995; McDaniel et al., 2004). PM performance is also influenced by certain characteristics of the ongoing activities, such as the level of complexity, difficulty and/or the cognitive resources necessary to perform the ongoing task (Einstein et al., 1997; Marsh, Hicks & Bink, 1998; Meier & Zimmermann, 2015).

Previous studies have proposed that successful self-initiated retrieval of the PM response depends on the strength of the association between the target cue (event) and the intended action, formed during encoding (Brewer, Knight, Marsh, & Unsworth, 2010) and on the event being properly processed as a cue at retrieval (Einstein et al., 1997). Consequently, significant efforts have been directed to identify encoding strategies to enhance prospective remembering.

One strategy, future thinking instructions, consists of instructing participants to imagine themselves performing an intention during encoding. This strategy draws from empirical evidence suggesting that simulating a certain action may increase the chances of performing that action in the future (e.g., Taylor & Schneider, 1989; Taylor, Pham, Rivkin, & Armor, 1998). Several experimental studies did find significant benefits of future thinking instructions on PM performance. Altgassen, Kretschmer and Schnitzspahn (2017) found that enhancing encoding of a PM task by asking participants to imagine themselves performing the task improved PM performance in both younger and older adults. Similarly, a future thinking manipulation during encoding benefited PM performance in a student sample (Neroni, Gamboz, & Brandimonte, 2014). Leitz et al. (2009) and Paraskevaides et al. (2010)

reported that future thinking instructions improved PM performance in young adults with acute alcohol consumption. The mechanisms underlying the effects of future thinking instructions on PM performance are still under debate. One hypothesis proposes that forming a visual representation of the future act may produce more durable memory traces of the intention by strengthening encoding. The alternative hypothesis is that future thinking encoding may produce a stronger cue-context association and thus elicit automatic retrieval of the intention (Paraskevaides et al., 2010). In line with the later hypothesis, Brewer, Knight, Marsh, and Unsworth (2010) have reported a strong positive correlation between the strength of the cue-to-context association formed at encoding and PM performance in an event-based task. Paraskevaides et al. (2010) reported that the benefits of future thinking instructions on PM performance are significantly larger for event-based compared to time-based PM tasks.

Altgassen, Kretschmer and Schnitzspahn (2017) compared PM performance under 3 encoding conditions (future thinking, repeated encoding and simple encoding) in adolescents, younger adults and older adults. Their results showed that future thinking instructions benefitted all age groups compared to simple instructions. Overall, PM performance was the highest under the repeated encoding condition, however, adolescents benefited more from future thinking instructions while younger adults from repeated encoding. Seen as repeated encoding enhanced PM performance beyond future thinking instruction, they propose that the most important mechanism underlying the effect of future thinking on PM performance is creating stronger memory traces of the PM intention.

Several researchers proposed that deep encoding of the cue-action association will automatically prompt the action /PM response upon occurrence of the cue (Einstein & McDaniel, 1990; McDaniel et al., 2004). Enhancing encoding may facilitate different components of a PM task: the retrospective component (recall of the intention - what is to be done), acknowledgement of an event as a cue (e.g., when I see the drug store) and remembering that you are supposed to something (Smith et al., 2014). One aim of the current study was to determine how these components are impacted by enhanced encoding over a long retention interval (1 month).

Another important question in PM literature is what contributes more to PM failures in event-based PM tasks: failure to acknowledge an event (context) as a cue for the intended action or failure to recall the action itself. Studies performed in laboratory settings (Einstein &

McDaniel, 1990) have established that the retrospective component of a PM task is significantly easier to remember than the PM component (Dismukes, 2010). In Einstein and McDaniel's (1990) experiment, participants who consistently failed to perform the PM task had, nevertheless, no problem in remembering when and what they were supposed to do (when asked at the end of the experiment). Therefore, the self-initiated/spontaneous retrieval of the PM task did not occur despite the fact that the association between the cue and the action to be performed has been encoded and stored in the memory. Previous research proposed that 'noticing' the prospective cue is more difficult and requires explicit memory but once the cue is noticed, searching the memory for the meaning of the cue (the action to perform in response to the cue) is relatively easy (Chasteen, Park, & Schwarz, 2001).

Most studies assessing different methods to stimulate successful PM task performance have been performed in laboratory setting. In laboratory settings the prospective step ('when') of the PM process typically happens over a relatively short interval. While in real life many PM activities cannot be performed so quickly, rather one often needs to form an intention that can only be carried out after days or weeks. Over long delays the relative contribution of retrospective and prospective components on successful completion of a PM task may change (Nigro & Cicogna, 2000; Dismukes, 2010), with the retrospective component becoming more relevant. In everyday life we often meet with situations in which we know that we are supposed to perform a certain action but cannot remember what the action is. Association with a former intent does not necessarily involve association with the specific action. A second aim of the current study was to identify whether the retrospective component of a PM task continues to be more difficult than the prospective component, even over long retention intervals (1 month).

There is no consensus in the literature regarding the effects of delay (retention interval) on PM task performance. Several studies found a decline in PM response at increased delays (Loftus, 1971; Brandimonte & Passolunghi, 1994), while others reported an increase in PM responses at longer delays (Hicks, Marsh, & Russell, 2000; Martin, Brown, & Hicks, 2011), and others still reported no influence of the retention intervals (Wilkins, 1976; Einstein, Holland, McDaniel & Guynn,1992; Guynn, McDaniel & Einstein, 1998). These studies, however, used laboratory paradigms and compared relatively short retention intervals. Nigro and Cicogna (2000) employed a more naturalistic PM task (remembering to give a message to a second experimenter) and found that the retention interval (10 minute, two days, two weeks) did not influence PM performance. Most studies which have compared PM performance in different task settings (i.e., laboratory, real life) have predominantly focused on age-related PM performance. Yet, even in these, the naturalistic tasks were more or less artificial, whether these findings generalise to actual real-life PM tasks remains to be determined (Schnitzspahn, Kvavilashvili & Altgassen, 2018; Meier, 2019).

The present study aimed to address the ‘when’ and ‘what’ steps of PM: by investigating PM performance over a longer term (1 month) and using more naturalistic experimental conditions to see if enhancing encoding, by repeated retrieval of the PM intention during encoding, will increase PM performance in the long-term as it does in the case of RM. We know that repetition typically helps with the long-term remembering of ‘what’ in classical episodic memory tasks. Given the association between episodic memory and PM, and the well-studied effects of retrieval on long-term memory, it was hypothesised that the repeated retrieval of the ‘what’ of the PM task (content of the intention) will aid participants in their performance. With regard to the ‘when’ the aim was to create a more realistic task, by evaluating PM performance over a long delay and in a different environment, by calling participants after 1 month. Lastly, long-term memory performance on a cued recall task over the same 1-month interval was measured and compared to long-term memory on the PM task.

Several authors have suggested that there is a relation between working memory and event-based PM and that this relation is mediated by individual differences in attention and episodic memory ability (Brewer, Knight, Marsh, & Unsworth, 2010). Present PM theories propose two strategies for the retrieval of PM intentions (Cohen & O'Reilly, 1996; Gollwitzer & Brandstätter, 1997; McDaniel & Einstein, 2000): intentions are kept active in working memory while monitoring the environment (Koechlin & Hyafil, 2007; Gilbert, 2011) or that intentions are stored in episodic memory and retrieved when the appropriate cue is encountered (McDaniel & Einstein, 2007; Beck et al., 2014). Yet, in instances where there is a long delay between the formation and possible execution of the intention, adopting a monitoring strategy would be too resource demanding, thus less likely. In naturalistic PM tasks, such as the one employed in the present study, where the delay between intention formation and the opportunity to execute the intention is long, it is unlikely that participants could sustain monitoring processes over the entire interval. It was hypothesised that in this case, the most likely PM strategy would be to store the intention in episodic memory, thus a relation between episodic RM and PM performance should emerge.

The appropriate way of assessing PM performance has raised much disagreement in the literature. Estimates of performance can either produce a single binary measure or multiple measure points of success or failure, however, both can introduce complications (Logie & Maylor, 2009; Maylor & Logie, 2010). Binary measures have been criticised for only being able to provide a rough estimate of PM ability. While using multiple measure points by increasing the number of PM target events, in an attempt to enhance precision, may change the nature of the task transforming it into one of vigilance (Uttl, 2005). Additionally, using multiple measures may increase the likelihood that a response on one PM target event will enhance the chance of responding to the subsequent PM target event (Maylor, 1996 cited in Logie & Maylor, 2009). While the present experiment avoids the problems that may arise when administering multiple PM trials, its limitation lies in the inclusion of only a modest sized group, resulting in limited statistical power.

6.2. Method.

6.2.1. Participants.

A total of 52 young participants were recruited mainly from Carol Davila University of Medicine and Pharmacy of Bucharest and Politehnica University of Bucharest, Romania. Participants were allocated to different degrees of encoding (higher or a lower condition): 1 repetition and 3 repetitions respectively of the PM task during encoding. Out of the 52 included, six participants dropped out, as they did not answer the phone for the 1-month assessment. The final sample included 46 participants, 22 in the higher encoding and 24 in the lower encoding condition.

Participants' written consent and demographic information concerning education, gender and age were obtained before starting the experiment. For the encoding phase participants were tested in person, for the remaining test delay participants were tested via telephone.

6.2.2. Material.

The material consisted of the same scrambled sentences used in Experiment 2 & 5 (Chapters 2 and 5, respectively; the full material can be found in the Appendices).

6.2.3. Design.

This experiment consisted of an encoding phase, where all participants were presented with the 16 sentences, these were read out by the experimenter (author of current thesis) at a slow and clear pace (2s pause between each sentence). To minimise any recency effects, each presentation phase was followed by a 1-minute filler task that involved creating as many words as possible from the Romanian word “hippopotam” (see Baddeley, Allen, Atkinson & Kemp, 2019). Participants then took the post-encoding cued recall test on one of the four subsets, which was self-paced.

Participants were then given an event-based PM task that required them to deliver the message ‘The sun is shining’ when the experimenter would call them at a later occasion for the follow up test. Participants were randomly assigned to either a higher encoding condition or a lower encoding condition. For each condition participants had to repeat the PM instruction (‘The sun is shining’) either three times or just once, in Romanian (the native language of all participants and experimenter).

The encoding phase and retrieval test were conducted face to face while the final test was conducted by telephone. This type of testing has been validated in previous studies (e.g., Baddeley, Rawlings & Hayes, 2014; Baddeley, Allen, Atkinson & Kemp, 2019; Stamate, Baddeley, Logie & Della Sala, 2020). All testing was conducted in Romanian (native language of this thesis’ author and all participants). After one month, each participant was called by the experimenter and tested on the sentence material (RM task), after which the experimenter engaged in a friendly 1- minute conversation to give the participant the opportunity to deliver the PM message. If the participant did not remember to deliver the PM response during this time interval, the experimenter gave them one or two prompts. The participants who did not remember to deliver the message, were asked towards the end of the call ‘Is there anything else?’ (first prompt); participants who could still not deliver the message were asked ‘Were you supposed to tell me anything else?’ (second prompt).

6.3. Results.

Two variables for the PM performance were constructed: one dichotomous variable where PM performance was scored as correct only if participants self-initiated the PM response (delivered the message without any prompt); one variable with 4 levels: self-initiated response; response after first prompt (remembered what/when to do, but failed to do it);

response after second prompt (remembered the content of the message only); no response. All participants remembered the PM content correctly.

For performance on the cued recall test three variables were constructed: individual score at post encoding assessment (the number of correct features recalled - RM task), individual score at 1-month assessment (RM); learning capacity (which classified participants in three categories according to their cued recall score at encoding: slower learners if the score was lower than 6, average learners if the score was 8, faster learners if the score was higher than 10).

The prospective component was more difficult than the retrospective component of the PM task: 54.3% of participants (25 out of 46 participants) performed the PM task correctly (spontaneously retrieved the event-based PM response). When adding the number of participants who remembered what action they were supposed to perform and when to perform it (retrospective component of the PM task) after first prompt (were asked by the experimenter: 'Is there something else?'), 95.7% of participants (44 out of 46 participants) were able to accurately remember the PM task.

A chi-square test of independence was run to examine the relation between PM encoding condition (high/low) and the ability to spontaneously retrieve the event-based PM response. The relation between these variables was significant [$X2$ $(1, N = 46) = 7.38$, $\varphi=.40$, $p = .007$]. Participants exposed to the high PM encoding condition were more likely to spontaneously retrieve the PM response compared to participants exposed to the low PM encoding condition (see Table 6.1).

A second chi-square test of independence was run to examine whether PM encoding condition (high/low) had any effect on memory for the retrospective component of the PM task [$X2$ $(1, N = 46) = 1.9177$, $\varphi=.20$, $p = .116$]. The relation between these variables was not significant. Participants in both PM encoding conditions were able to remember the retrospective component of the PM task.

Group	Low degrees of learning condition		High degrees of learning condition	
	N	% of N	N	% of N
correct PM response (self-initiated)	9	37.5	17	77.3
no PM response	15	62.5	5	22.7
PM response after 1st prompt (Is there something else?)	13	54.2	5	22.7
PM response after 2nd prompt (Where you supposed to tell me something?)	2	8.3		

Table 6.1. Number and percentages of participants falling into different PM response categories as a function of encoding condition.
N: number of participants.

Further analyses were carried out to see whether recall performance on the story recall task at post encoding assessment, recall performance on the story recall task at 1-month assessment and PM encoding condition are significant predictors of event-based PM spontaneous retrieval. A binary logistic regression was run to understand whether PM response (dependent variable measured on a dichotomous scale – "yes" or "no") can be predicted based on recall performance on post encoding assessment, recall performance at 1 month (independent continuous variables measured as the total number of correct responses). Logistic regression results indicate that neither post encoding recall performance ($p= .170$) nor recall performance at 1 month ($p= .871$) were significant predictors of PM response. Overall recall performance explains very little (6%) of the variance in the ability to provide the PM response.

It was further investigated if learning capacity (independent categorical variable with 2 levels: slower learners, faster learners) can predict PM performance, Chi-square tests were run separately in each encoding condition, these showed that the ability to spontaneously retrieve the event-based PM response was not related to learning capacity (low encoding condition: [X2 (1, N = 23) = .014, φ=.024, p= .907]; high encoding condition: [X2 (1, N = 22) = 3.115 φ=3.76, p = .078]).

6.4. Discussion.

The current results show that enhancing encoding by repeating the PM response during encoding did facilitate self-initiated retrieval of the PM response at one month. Participants in the higher encoding condition were significantly more likely to self-initiate the PM response (deliver a message to the experimenter) compared to participants in the lower encoding condition. Previous studies investigating implementation intentions encoding in the context of goal-oriented behaviour showed that implementation intentions encoding had a medium to large (for a review see Gollwitzer & Sheeran, 2006) and durable effect on successful goal completion (Papies, Aarts & de Vries, 2009). These results show that enhancing encoding had a significant effect over PM performance over a long retention interval.

Another finding from this experiment is that over long intervals (1 month) the prospective component of a PM task (self-initiated retrieval of the intention/PM response) is still more difficult compared to the retrospective component (remembering what the intention is and when to perform it). Irrespective of encoding condition, 43 out of 46 participants (97%) successfully stored and retained the association between the prospective cue and the intended action (remembered what they were supposed to do and when) over the 1-month interval. Only 25 of them were able to successfully self-initiate the PM response (to deliver the message without a prompt from the experimenter). These results reinforce the fact that a strong and long-lasting association between the cue and the intention does not necessarily insure a successful self-initiated PM response. Similar results were reported by Nigro and Cicogna (2000). In their experiment failure to deliver the message (lack of PM response) accounted for 42.86% of the PM errors, whiles errors of the retrospective memory component (‘errors what’- wrong content of the message, ‘errors when’ - message delivered at the wrong time, and ‘errors what and when’- incorrect message delivered at the wrong time) accounted for 27.18% of the PM errors over a 2 weeks retention interval.

Thus, in the current study PM failures occurred due to participants failing to acknowledge the event (receiving a call by the experimenter) as a cue for the PM response (to deliver the message), and not due to participants not being able to remember what action they were supposed to perform or, as a matter of fact, when to perform it.

In this experiment enhanced encoding only facilitated the prospective component of the PM task and had no effect on the retrospective component. This result may be due to the fact that the retrospective component of the current task was relatively easy (all participants in the high degrees of learning condition, and all but 3 in the low degrees of learning condition were able to remember the retrospective component correctly). One study performed in laboratory settings (Smith et al., 2014) and employing a PM task which was more difficult and complex, in terms of both the prospective and retrospective component, but much shorter in delay interval, found that implementation intentions encoding enhanced both self-initiated response to a target cue and memory for the content of the action (remembering what needs to be done). One previous study directly investigating the effect of delay on PM performance (Nigro & Cicogna, 2000) reported that the retrospective component is stable over long retention intervals. They found that memory for the 'what' component of the PM task was relatively similar at 10 minutes, 2 days and 2 weeks, even though their retrospective component of the PM task was more difficult than in my experiment.

PM performance was not related to the RM cued recall performance. One explanation for this lack of correlation could be explained by the levels of difficulty of the tasks: the cued recall task was difficult while both the retrospective and prospective components of the PM task were relatively cognitively undemanding. Einstein and McDaniel (1990) proposed that PM performance might be influenced by retrospective memory in PM tasks where the retrospective component has a high level of difficulty and complexity. Accordingly, when a PM measure has a heavy retrospective memory component (e.g., when participants have to respond to 30 different PM cues, that is, remember 30 cue-action pairs), successful performance will depend more upon the retrospective rather than the prospective component, and the index will have low validity in measuring PM ability. The results of Experiment 6 are in line with previous research consistently reporting no correlation between RM and PM (Wilkens & Baddeley, 1978; Meacham & Leiman, 1982; Einstein & McDaniel, 1990; Shelton et al., 2016).

6.4.1. Conclusion.

Using a naturalistic task, Experiment 6 provided evidence that enhancing encoding of a PM task can positively influence event-based PM performance over long intervals. The findings show that the benefits of enhanced encoding, reported in laboratory experiments can extend to more everyday-like situations. Furthermore, the relative contribution of the retrospective versus the prospective component does not change when assessed over longer retention intervals or in everyday life contexts. Even after long retention intervals, self-initiated retrieval of an intention (the prospective component of the PM task) is still more difficult compared to the retrospective memory component of the PM task (remembering the intention/action to be performed and when to perform it). A third finding was that correctly recalling the association between the cue and the intention does not warrant self-initiated retrieval of the intention. With the current manipulation, enhancing encoding by having participants repeat the PM memory task once or three times did enhance self-initiated retrieval without having any significant effect on the recall of the PM task (what and when). Fourth, even when the PM task is performed a long time after the intention was encoded, PM performance does not depend on RM capacity. One practical implication of the current findings is that the simple act of repeating an intention/response several times during encoding can successfully increase the chance of performing it.

CHAPTER 7: Forgetting in routine everyday tasks - habitual prospective memory.

The initiation of action, in everyday life, relies on accurate remembering of past events, even those events that were encountered once and processed incidentally (Cubelli, 2010). Thus, the encoding phase of the memory process, which seldom includes intentional learning (outside the laboratory), is also involved in forgetting. When faced with such events, we are able to remember the gist of that event, but we cannot encode all relevant and irrelevant details pertaining to it (Clifasefi, Garry, & Loftus, 2007). Therefore, we can later remember only what we have encoded, and we encode based on purpose and actual knowledge, with other information being forgotten (Cubelli, 2010).

The encoding phase for events that are encountered repeatedly can also be processed incidentally, with much of the rich contextual details being forgotten. In several instances, forgetting of these details from the encoding phase, can cause significant consequences. For example, the ability to remember whether we medicated today or just the day before, can pose a significant threat to our health. When having to perform an intention in the context of these routine everyday tasks, which refer to habitual prospective memory (PM), the necessity of initiating (or not) a certain action is highly dependent on the accurate memory of the previously performed action (Marsh et al., 2007; McDaniel et al., 2009).

PM involves a process of reconstruction, it is sensitive to interference and requires reminder cues to be associated with a previously established intention (Guynn, McDaniel, & Einstein, 1998). True recognition of a previously performed action/task requires memory for detailed contextual information, while false recognition is based on a feeling of “déjà vu” (Brainerd & Reyna, 2002). Thus, significant attention has been given to the encoding phase of PM tasks in an attempt to reduce interference and promote the encoding of detailed contextual information.

The following experiment investigated whether enhancing encoding during the execution of a PM habitual action, by using an external stimulus, increases performance memorability and thus leads to successful completion of the intention, and examined the underlying mechanisms of forgetting in a habitual PM task.

Experiment 7: The benefits of enhanced encoding in habitual prospective memory

7.1. Introduction.

The ability to remember to perform a previously formed intention at the appropriate time has been defined as PM (Brandimonte, Einstein, & McDaniel, 1996; Ellis & Kvavilashvili, 2000; McDaniel & Einstein, 2007; Kliegel, McDaniel, & Einstein, 2008). When these actions involve routine everyday tasks, such as remembering to take a medication every morning or brushing our teeth several times a day, they are referred to as habitual PM (Meacham & Leiman, 1982 cited in, Hicks, Marsh, & Russell, 2000). Frequent performance of a PM task changes it from episodic to habitual.

Apart from execution frequency, another clear difference between episodic and habitual tasks relates to intention formation. Each time we plan an episodic PM task, such as paying a bill or sending a letter, we form an explicit intention, whereas in habitual PM tasks, intention to perform an action is often implicit. We do not explicitly form the intention to switch off the gas after we prepared coffee or to lock the door when we leave home (Meacham & Leiman, 1982 cited in, Hicks, Marsh, & Russell, 2000). It has been proposed that these two types of memory can also be distinguished in terms of the neural networks that support them (Meier et al., 2014).

Experiment 7 investigated whether enhancing encoding during the execution of a PM habitual action, by using an external stimulus, increases performance memorability and therefore decreases repetition errors (REs), and also examined the underlying mechanisms of forgetting in a habitual PM task.

7.1.1. PM failures.

There are two ways in which PM can fail: omission errors (OEs) such as forgetting to complete an action (e.g., forgetting to take a pill) or REs by erroneously repeating the completed action (e.g., taking the same pill twice).

The literature suggests that episodic PM tasks are more prone to OEs while habitual PM tasks are more susceptible to REs as they are more likely to create ambiguity regarding whether or not the task has already been performed. Einstein, McDaniel, Smith, and Shaw (1998) found that in multiple repeated trials of the same task, the pattern of errors changes across trials; on early trials, when the task is still novel, OEs are more frequent, whereas as the task becomes habitual participants make more REs.

Another important variable responsible for the pattern of PM errors is age. Studies have generally found that younger participants are more likely to overestimate their performance, leading to more OEs while older participants underestimate their performance and are more likely to repeat the same task (Marsh et al., 2007). This is especially true in attentionally demanding situations, whereby older participants have difficulties in remembering whether a habitual action has actually been performed and, when unsure, they tend to repeat the action rather than omit it (McDaniel et al., 2009).

7.1.2. Retrieval in PM.

While OEs have been frequently investigated (Kvavilashvili, 1992; Kliegel, Martin, McDaniel, & Einstein, 2001) the interest in REs is more recent. Existing studies have mostly examined factors that lead to REs at retrieval, yet only a few studies have investigated the impact of encoding strength during PM actions.

Pink and Dodson (2013) found that in subsequent tasks requiring divided attention, participants who responded to habitual PM cues in a previous task were significantly more prone to REs than participants who responded to infrequent PM cues.

7.1.3. Encoding in PM.

Several studies found that implementation intention encoding (IIE), an encoding strategy aiming at strengthening the association between a situational cue and an intended action (Gollwitzer, 1999), can be very efficient in reducing OEs in a variety of habitual or episodic

PM tasks (Orbell, Hodgkins & Sheeran, 1997; Sheeran & Orbell, 1999; Liu & Park, 2004, cited in McDaniel & Scullin, 2010). IIE has been shown to improve performance of a PM tasks in young participants (McDaniel et al., 2008; McDaniel & Scullin, 2010; Meeks & Marsh, 2010, Brewer et al., 2011; cited in Chen et al., 2015) as well as in older participants (Chasteen, Park, & Schwarz, 2001; Liu & Park, 2004; McFarland & Glisky, 2011, Burkard et al., 2014, cited in Chen et al., 2015). While the positive effect of IIE on PM performance is largely acknowledged, the underlying mechanisms are still unclear.

Gollwitzer (1999) proposed that IIE produces high activation and accessibility of the situational cue and prompts automatic performance of the PM task when the situational cue is presented (Gollwitzer, 1999; Bayer, Achtziger, Gollwitzer, & Moskowitz, 2009, cited in McDaniel & Scullin, 2010). Other studies (McDaniel & Scullin, 2010; Chen et al., 2014) propose that while IIE does benefit PM performance by strengthening the association between the cue and the intended action, it does not support a completely automatic PM response. Performance of a PM task under IIE may involve both automatic and controlled processes: whereas occurrence of the situational cue may prompt spontaneous remembering of the PM intention, the actual execution of the PM task requires cognitive resources, such as switching attention from the ongoing task to the PM task.

Though enhancing encoding, by strengthening the association between an environmental cue and an intended action, has proven effective in reducing OEs, it may have the opposite effect on REs. Bugg, Scullin and McDaniel (2013) found that stronger encoding of the intention-cue association led to more REs as it prompted spontaneous retrieval even when the intention was no longer relevant in the current task.

Remembering if a task/action has or has not been performed is particularly challenging in habitual PM tasks (Einstein, McDaniel, Smith, & Shaw, 1998). This has been attributed to a failure of different mental experiences, poor internal-source monitoring (confusing the action of taking medication with the thought about taking the medication; Johnson, Raye, & Estes, 1981) and temporal discrimination (Friedman, 1993). Older individuals are especially more prone to source confusions (Thomas & Bulevich, 2006; McDaniel, Lyle, Butler, & Dornburg, 2008). These, in turn result in more REs compared to OEs (Hashtroudi, Johnson, & Chrosniak, 1989).

Marsh et al. (2007) conducted two consecutive event-based PM experiments with younger and older participants. In order to facilitate encoding of the PM task, they instructed participants to additionally turn to the investigator and name aloud the target event whenever they pressed a designated PM key. They found that increasing PM task complexity benefitted younger participants but had the opposite effect on older participants. They proposed that this difference was due to the task instructions, which may have generated retrieval competition, increasing older participants' confusion regarding to which event they responded and to which they did not, especially since the events were semantically related. On the contrary, McDaniel et al. (2010) found that a more complex motor action (putting one hand on one's head, while pressing the designated PM key) increased performance memorability and was effective in reducing REs in older participants. They proposed two possible explanations: that performing a more complex action either engaged more attentional resources, thus enhancing output monitoring; or it provided more sensory information, thus increasing the participant's confidence that they have performed the action.

7.1.4. Recollection and familiarity in habitual PM.

A recent study by Sadeh, Ozubko, Winocur and Moscovitch (2014) proposed that the way we forget may depend on how we remember, the characteristics of forgetting being influenced by the underlying declarative memory representations: recollection or familiarity. Recollection involves conscious awareness of an event and its contextual information, whereas familiarity involves a feeling of "deja vu", void of any contextual information.

In habitual tasks, the necessity of initiating (or not) a certain action is highly dependent on the accurate memory of the previously performed action (Marsh et al., 2007; McDaniel et al., 2009). True recognition of a previously performed action/task requires memory for detailed contextual information, while false recognition is based on a feeling of "déjà vu" (Brainerd & Reyna, 2002). This distinction was also noted at a neuroanatomical level. Functional neuroimaging studies (Kim & Cabeza, 2007) found that high confidence in true recognition engages mostly medial temporal lobe regions associated with recollection, while high confidence in false recognition engages mostly the frontoparietal regions, associated with familiarity. Meier and colleagues (2014) proposed that frequent retrieval of a task may cause a change in parietal old/new effect activity. Several previous studies have shown that the parietal old/new effect is enhanced when confidence on recognition is higher being based on ‘remembering’ rather than ‘knowing’ (Smith, 1993; Düzel et al., 1997; Schloerscheidt et al., 1998; Trott et al., 1999, cited in Curran, 2004).

Experiment 7 investigates the hypothesis that individuals forget the execution of habitual tasks because of the way the memories of these completed actions are encoded. When we repeatedly execute a task it almost becomes automatic and we no longer pay attention to what we are doing, therefore we encode far less contextual information. Presumably, increases in habitual PM REs are primarily based on familiarity-related activity, whereas high confidence in remembering the executed task is likely to be associated with recollection. Cheyne, Carriere and Smilek (2006) suggested that even though everyday memory errors have multiple determinants they are primarily attentional in origin, and that at least part of these problems are a consequence of lapses of attention at the time of encoding. The idea that attention failures lead to memory failures is consistent with numerous studies showing that dividing attention decreases memory performance and allows retrieval to be dominated by automatic processes (e.g., familiarity) rather than conscious recollection (Jennings & Jacoby, 1993; Smallwood et al., 2003). The way we attend to stimuli at the time of study has an important influence on familiarity/recollection (Johnson, Raye, & Estes, 1981). Since encoding the features of an action and its context facilitates both internal-source monitoring (Johnson, Raye, & Estes, 1981) and temporal discrimination processes, dividing attention should account for more PM errors in habitual prospective remembering than in episodic prospective remembering.

It was predicted that any additional information provided during the encoding of the PM task will result in a more distinctive memory record. Previous studies have suggested that the reason older adults have a relatively poor memory for their past PM performance may be due to an encoding deficit (Marsh et al., 2007). This is also the case with ‘poorer’ measures of output monitoring that require younger and older adults to free recall lists of words and later judge from all of the studied items which they had previously recalled and which they had not (Koriat, Ben-Zur & Sheffer, 1988). Older adults tend not to identify all the words they had correctly recalled suggesting that items do not become bound to the free recall act in the same way as they do for younger adults (i.e., an encoding difference). Studies of ageing and self-performed tasks (SPTs), tasks where participants perform simple actions such as spin the top, found that with less engaging actions older adults had worse memory for SPTs than did younger adults, while with more engaging activities the age-related decline disappeared (Kausler & Hakami, 1983).

Based on the previously mentioned research, the following predictions were made for the current study:

1. Participants will show increased REs on habitual PM tasks when the encoding of the completed action is performed automatically, indicating that this automatic encoding is mediated by familiarity processes.
2. When encoding is strengthened with the use of an additional cue, to provide additional contextual details, participants will show fewer REs.
3. Subjective memory performance with regard to both PM and RM components, as measured with the Retrospective Memory Questionnaire (PRMQ; Smith, Della Sala, Logie, & Maylor, 2000), will not be reflected by objective memory performance measured with the laboratory experiment.

7.2. Method.

From extant literature I could glean two studies investigating enhanced encoding conditions for a PM task, one by McDaniel et al. (2009) and the other by Marsh et al. (2007); they report contrasting results. McDaniel et al. (2009) found that older adults' REs were significantly reduced when they were required to perform a distinct motor activity. On the contrary, in the study by Marsh et al. (2007) the manipulation in the enhanced condition created retrieval competition in older adults, whose performance hence decreased.

The current study was devised to reconcile these idiosyncrasies. To this end, the paradigm based on Einstein et al. (1998)'s procedure was modified as follows: Instead of asking participants to perform a complex motor action to provide additional source information of the PM task, this was incorporated within the task: the words 'You clicked!' appeared briefly on the screen each time participants pressed the designated PM key. Secondly, a within-subject design was used. Instead of allocating participants to either the enhanced encoding condition or the standard one, in the current design these were intermixed within the same experiment. This modification is beneficial not only because the variability in measurements is more likely due to differences among conditions than to behavioural differences across participants (relevant if a between-subjects design were used), but mainly it enabled the comparison of remember (R) and know (K) judgements. The remember-know paradigm has been used by several researchers to investigate issues regarding memory awareness (Hay & Jacoby, 1999). Participants are asked to subjectively report their memory of certain stimuli basing their classification on whether they ‘remember’ (they recall a specific detail about the stimuli’s previous occurrence) or ‘know’ (they cannot remember any details of its prior presentation) stimuli that have been previously presented.

The final change to the original design was to ask participants how confident they are that they will remember having performed the PM action during each trial (12 trials). This was intended to reveal the processes within and between the PM blocks (enhanced vs. standard encoding). It was hypothesised that there would be more R judgements for the enhanced encoding condition and K for the standard. It was further predicted that the responses based on R judgements will produce less REs. Although the Remember-Know procedure is the most commonly used to assess recollection and familiarity, it has also received a lot of criticism, specifically with regard to whether or not participants actually understand the distinction between R and K judgements. In an attempt to address this issue, a confidence scale was used to assess familiarity and recollection (e.g. How confident they are on a scale from 1-5 that they will remember having pressed the PM action key? 1-Very sure not; 2-Quite sure not; 3- Don’t know; 4- Quite sure I will; 5- Very sure I will) after each PM task.

Other than the modifications discussed above the procedure closely followed that of Einstein et al. (1998).

7.2.1. Procedure.

The entire experiment was computer-based, conducted in PsychoPy (see Figure 7.1. in the Appendices). Participants were told that the experiment was measuring their ability to perform a number of tasks, as well as their ability to remember to perform future actions (PM task). Specifically, they were asked to remember to press a designated key (mouse click - PM action) on the keyboard during each of 12 trials (2 min each). They were instructed not to press the key immediately after the trial started but to wait for approximately 30 seconds, self-timed (however, they were not allowed to use a watch).

The experiment consisted of 2 parts, namely, 5 practice trials and the 12 main trials. The practice trials had the same design and instructions as the main trials (an ongoing task + PM task), however, for the practice trials the ongoing task was always the same (a letter task). Participants who were unable to complete the practice trials did not go on to perform the main experiment. For the main experiment's trials, participants were tested on different tasks (see Table 7.1 describing the individual tasks), spaced out over the 12 trials. Each of the tasks was designed to measure a different construct or cognitive ability, lasted for approximately 2 minutes and had a different background colour (to increase memorability). The nature of the tasks, as well as the required responses were described on an introduction screen before the start of each trial. The standard and enhanced condition was intermixed between trials, with the condition sequence allocated randomly across participants (e.g. odd number participants started with the enhanced condition). The enhanced condition consisted of the message "You clicked" together with a drawing of an animal (different drawing for each trial) appearing after participants performed the PM action (mouse click). After the message and image disappeared, a screen with the R/K confidence scale was displayed. After participants inputted their answers the ongoing task continued. This procedure was the same for all 6 of the enhanced trials. While for the standard trials, no messages or R/K confidence scale was displayed.

Ongoing task description
Trial 1. Vocabulary - word meaning task, deciding between real and made up words.
Trial 2. Perceptual Speed - identify the number letters that appear in pairs.
Trial 3. Word recognition - spot the odd word (3 words, 2 from the same semantic category).
Trial 4. Action control questionnaire.
Trial 5. Processing speed - Stroop task.
Trial 6. Speed and processing task - measures participants ability to manipulate numbers and make simple arithmetic computations.
Trial 7. Reasoning - measures participants ability to think logically and solve abstract problems.

Trial 8. Implicit memory - word fragment completion test measures memory of words participants have merely been exposed to.
Trial 9. PRMQ questionnaire.
Trial 10. Spatial visualisation - assesses participants ability to think spatially and mentally manipulate images and perceive patterns between them.
Trial 11. GDS
Trial 12. Word pleasantness rating - self-reported pleasantness rating for different words (scale 1-5).

Table 7.1. Description of the ongoing tasks for each of the 12 trials in the experiment.
PRMQ: Prospective Retrospective Memory Questionnaire (Smith, Della Sala, Logie, & Maylor, 2000); GDS: The Geriatric Depression Scale (Yesavage, et al., 1983); Stroop task (Stroop, 1935).

After completing all 12 trials participants had to answer another confidence scale regarding their RM of the PM performance. Specifically, participants were asked to rate (on a scale from 1-5) on how confident they are they have performed the PM task on each trial, print screens of the trials were displayed together with the questionnaire (to get participants to think back to the PM task on each of these tasks). This enabled the comparison of the R/K encoding judgements done during the PM task to the answers participants gave on this last confidence judgement questionnaire.

7.2.2. Participants and design.

A total of 60 participants (30 older, 30 younger) were recruited for the study. The younger participants were recruited mainly among friends, acquaintances and some from the City Halls' Club in Brasov, Romania. The older participants were all healthy, community-dwelling adults recruited from the City Hall Seniors' Club and by GPs. Participants' written consent and demographic information concerning education, self-report of health and medication, and past or present medical conditions, were obtained before starting the experiment. Besides taking part in the PM experiment, all participants performed Raven's Progressive Matrices IQ test (Raven, 1962) and the elders were also tested with the MMSE (Folstein, Folstein & McHugh, 1975).

Four participants dropped out of the test before completion (3 older, 1 younger), a further 6 had to be removed from the analyses: two participants had to be excluded because of their performance on the practice trials, three participants were excluded due to incomplete data (PsychoPy error/malfunction), 1 younger participant due to (unreasonably low) performance on the IQ test.

The final sample included in the analysis consisted of 23 younger and 27 older participants (N=50). The younger participants had a mean age of 25.17 years (SD=3.49), average education 15 years (SD =1.93), a mean IQ of 115 (SD=10.04) with 56% of the sample population being women. The older participants had a mean age of 66.14 years (SD=5.35), average education 14 years (SD=2.73) with 85% of the sample population being women. Because IQ scores differed across younger and older participants, correlations between IQ scores and REs, OEs and RM for each group were measured. No significant correlations emerged.

Each participant was tested individually in sessions that lasted approximately 1 hour for younger participants (40 min the computer experiment and 20 min the IQ test) and approximately 1h and 40 minutes for older participants. The difference in time between groups is due to the fact that older participants typically took an average of 30 minutes more on the computer task (because of their limited computer skills) and because they were also tested on the MMSE.

The design was a 2 by 2 mixed factorial design, in which the experimental variables (enhanced PM/standard PM) and group (younger participants, older participants) were manipulated.

7.2.3. Measurement variables.

PM performance – performance on the PM task was measured as the mean proportion of trials where each participant pressed the action key (mouse click) only once across the 12 trials (trials where participants performed a repetition error were excluded).

Repetition errors (REs) – repeating the PM task (pressing the mouse button more than once) within the same trial was considered a REs. Unless otherwise specified in the results section, REs were measured as the mean proportion of trials where each participant pressed the response key more than once across the 12 trials.

Omission errors (OEs) – failure to perform the PM task within a trial was scored as an OEs. OEs were measured as the mean proportion of trials where each participant omitted to press the response key across the 12 trials.

Retrospective memory (RM) – based on confidence judgement scores related to past performance, rated in the post experiment questionnaire, a categorical variable for the accuracy of the RM for the past performance, for each trial: 1=correct, 0 = incorrect (irrespective of the confidence judgement ratings). Trials with REs were considered as incorrect (score=0). RM was measured as the mean proportion of trials where participants remembered correctly whether they had performed the PM task or not.

RM failures – not performing the PM task during a trial and indicating incorrectly on the post experiment questionnaire that the task had been performed and performing the PM task during a trial and indicating on the post experiment questionnaire that the task had not been performed (irrespective of the confidence level) were considered an RM failure. A categorical variable for RM failures was created, for each trial: 1=RM failure, 0 =correct. Trials with REs were considered as RM failure (score=1). RM failure was measured as the mean proportion of trials where participants remembered incorrectly whether they had performed the PM task or not.

RM confidence ratings – were based on confidence scores related to past performance, rated in the post experiment questionnaire, on a scale from 1 (very sure I did not click) to 5 (I am very sure I did click). Scores 1 and 5 were associated with recollection, scores 4 and 2 with familiarity; score 3 was I don't know.

The RM scores for each trial were measured as follows:
If PM was performed and the answer provided in the questionnaire was: 5 (I am very sure I did click) then the score will be 5; 4 (I think I clicked) the score will be 4; 3 (I don't know) score will be 3; 2 (I thinks I did not click) score will be 2; 1 (I am very sure I did not click) score will be 1.

If the PM was not performed and the answer to questionnaire was: 5 (I am very sure I did click) then the score will be 1; 4 (I think I clicked) the score will be 2; 3 (I don't know) score will be 3; 2 (I think I did not click) score will be 4; 1 (I am very sure I did not click) score will be 5.

In order to compare performance patterns as the task became more repetitive, the 12 trials were divided into 3 trial blocks. This allowed for a within-subjects variable (of trial-block) for each of the variables described above to be created.

7.3. Results.

7.3.1. Prospective memory (PM) performance.

PM performance was on the whole better for younger participants (M=.83, SD=.37) then for older participants (M=.558, SD =.50). One-way analysis of variance (ANOVA) showed a statistically significant effect of group on the total number of trials where the PM task was performed correctly [F (1, 49) = 39.214, p<.001].

PM performance was relatively stable across the 3 trial blocks in both groups. There was a slight decrease in performance from trial block 1 (younger M=.86, SD=.35; older M=.57, SD=.52) to trial block 2 (younger M=.77, SD=.42; older M=.52, SD=.50) followed by a slight increase in trial block 3 (younger M=.88, SD=.33; older M=.58, SD=.51) but neither the main effect of trial block nor the interaction effects between trial blocks and the age group approached statistical significance.

7.3.2. Repetition errors (REs).

The mean proportion of trials where the older participants committed REs was higher compared to younger participants. The assumption of homogeneity of variance was violated; therefore Brown-Forsythe F-ratio is reported; there was a significant effect of group on the mean proportion of REs made across trials [F (1,42.624) = 11.666, p<.001].

7.3.3. REs by condition.

As hypothesised, the number of REs made by the younger adults and the older adults differed as a function of condition. Participants in both groups were more prone to make REs in the standard condition (where no cue appeared after completion of the PM action) compared to the enhanced condition (for this analysis trials with no PM task were filtered out). Mixed analysis of variance (ANOVA) with the between-subjects variable set as group and within-subjects variable as condition (standard, enhanced) showed a significant effect of condition [$F(1,48) = 9.683$, $p = .003$] and a significant effect of group [$F(1,48) = 145.864$, $p < .001$] on the mean proportion of REs. There was no statistically significant group by condition interaction. Paired samples t-test showed that the differences between the two conditions were significant in each group (younger: enhanced condition $M = .0145$, $SD = .048$; standard condition: $M = .06$, $SD = .12$) $t(22) = -2.336$, $p = .029$; older: enhanced condition: $M = .119$ ($SD = .133$) standard condition: $M = .267$ ($SD = .285$) $t(26) = -2.689$ $p = .012$).

Though the enhanced condition was effective in significantly reducing REs, this effect was lost by the end of the experiment. Later in the results it is seen that the manipulation in the enhanced condition did not improve RM for the PM performance. The mean proportion of REs when participants declared they were very sure (recollection) they had performed the PM task was compared with the mean proportion of REs at lower levels of confidence (familiarity) and a linear model of REs as a function of level of confidence was constructed. The main effect of level of confidence was again not significant [$F(1,226) = 2.532$, $p = .113$].

7.3.4. REs by trial block.

A 3 by 2 mixed analysis of variance (ANOVA) with group as between-subjects variable and trial block (1,2,3) as within-subjects variables showed no statistically significant effect of trial block [$F(2,96) = 2.340$, $p = .102$] and a significant effect on group [$F(1,48) = 10.752$, $p = .002$] on the mean proportion of REs. The interaction between group and trial block was not statistically significant. Paired samples t-test showed that the mean proportion REs committed by the younger adults did not significantly differ across the 3 trail blocks. In the older group there was a significant decrease in the mean proportion of REs errors from trial block 1 to trial block 2 [$t(26) = 2.237$, $p = .025$] and a significant increase from trial block 2 to trial block 3 [$t(26) = -2.087$, $p = .048$ - see Figure 7.2].

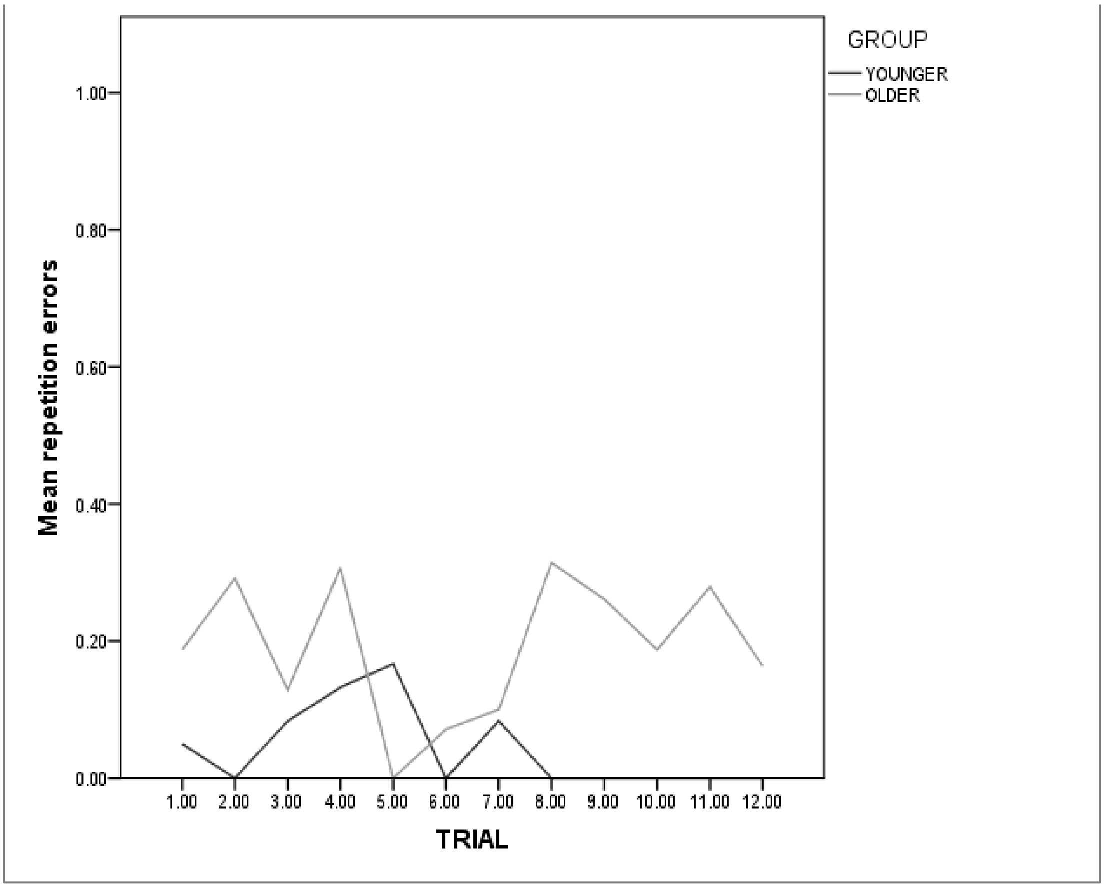


Figure 7.2. Mean proportion of repetition errors per trial for the younger and older adults (of the total number of trials with PM task).

7.3.5. Omission errors (OEs).

The mean proportion of trials where the older group failed to perform the PM task (M=.299, SD=.18) was higher than for the younger (M=.1268, SD=.10), similar to Einstein et al.'s (1998) study where older participants committed twice as many OEs compared to younger participants. The assumption of homogeneity of variance was violated, therefore Brown-Forsythe F-ratio is reported: there was a significant effect of group on the mean proportion of OE [F (1, 48) =17.369, $p<.001$].

7.3.6. OEs by trial block.

The number of OEs differed across the 3 trail blocks for both groups (see Table 7.2. and Figure 7.3). A 3 by 2 mixed analysis of variance (ANOVA) with group as between-subjects variable and trial block (1, 2, 3) as within-subjects variable showed a statistically significant effect of trial block [F (2,96) = 6.093, p=.004] and a significant effect of group [F (1,48) = 15.953, $p<.001$] on the mean proportion of OEs. The interaction between group and trial block was not statistically significant. Post hoc pairwise comparisons with Bonferroni correction showed that there was a statistically significant decrease in OEs from trail block 1 to trail block 2 (M= -138; p=.003). The mean proportion of OEs increased again in trail block 3 but the difference was not statistically significant.

Trial Block	**Prospective memory task performance**	**Younger**	**(N=23)**	**Older**	**(N=27)**
Trial block 1	Correct PM task	0.86	(0.35)	0.57	(0.52)
Trial block 2	Correct PM task	0.77	(0.42)	0.52	(0.50)
Trial block 3	Correct PM task	0.88	(0.33)	0.58	(0.51)
Trial block 1	OEs	0.08	(0.27)	0.24	(0.43)
Trial block 2	OEs	0.18	(0.39)	0.41	(0.49)
Trial block 3	OEs	0.12	(0.33)	0.25	(0.44)
Trial block 1	REs	0.07	(0.25)	0.19	(0.41)
Trial block 2	REs	0.04	(0.21)	0.07	(0.26)
Trial block 3	REs	0.00	(0.00)	0.17	(0.40)

Table 7.2. Mean proportion and standard deviation of correct PM task responses, OEs and REs as a function of age and trial block.

PM: Prospective Memory; OEs: Omission Errors; REs: Repetition Errors; N: number of participants.

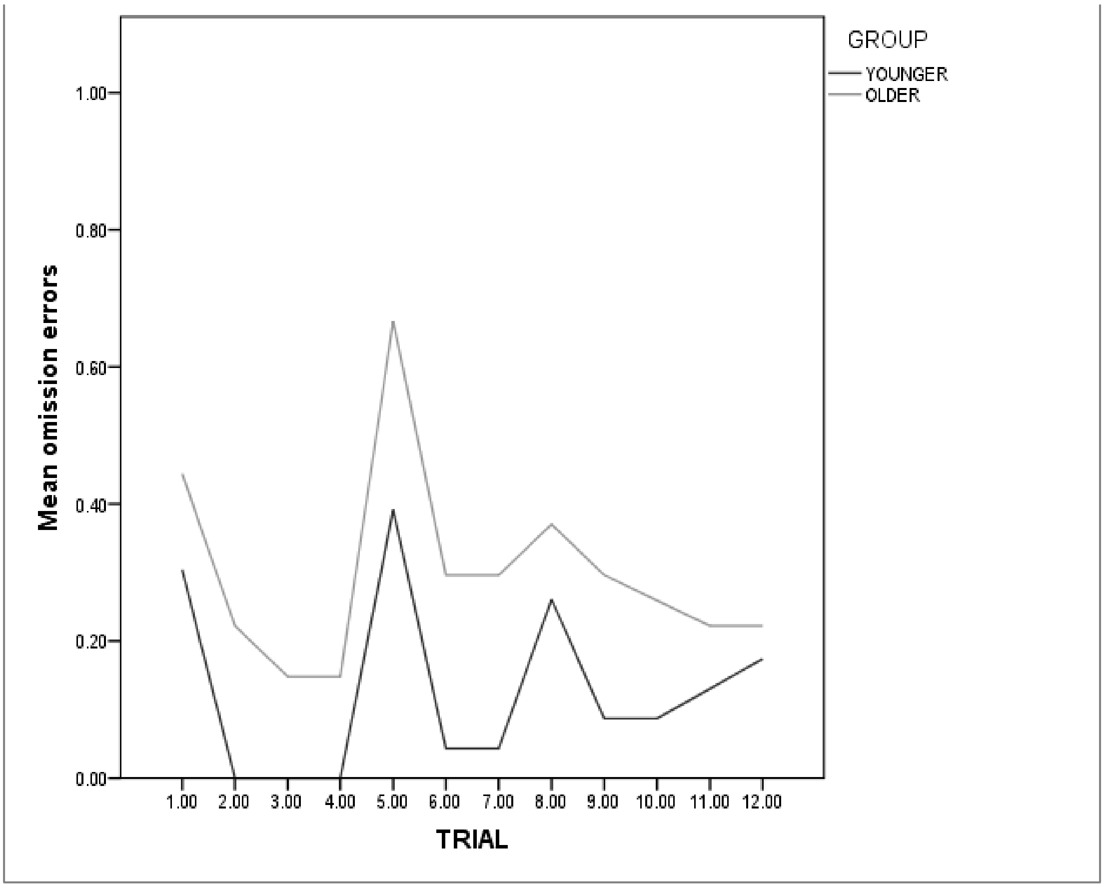


Figure 7.3. Mean proportion of omission errors per trial for the younger and older adults.

The significant increase in OEs in trial block 2 in this experiment may be due to the fact that trial 5 (Stroop test) and 8 (implicit memory test) were more attentionally demanding. The linear model of OEs as a function of trial showed a statistically significant effect of trial [F (1,596) =13.200, p<001] on the total number of OEs but no significant effect of trial by group interaction. It is worth mentioning that 35% (46 out of 132) of the total number of OEs were acknowledged by the participants (19 by younger, 27 by older participants) who answered in the post experiment debriefing that they were very sure (34 instances) or that they thought (12 instances) they had not performed the PM task. This aspect will be further discussed later in the results section.

7.3.7. PM errors by type: REs vs OEs.

The two types of PM errors (REs vs OEs) were compared, both within the same group and between groups. Older participants made significantly more REs as well as OEs compared to younger participants. Paired samples t-test showed that the mean proportion of OEs was significantly higher than the mean proportion of REs in both younger group and older group (younger group: REs $M=.036$, $SD=.07$; OEs $M=.1268$, $SD=.1$, $t=-2.969$, $p=.007$; older group: REs $M=.14$, $SD=.13$; OEs $M=.29$, $SD=.18$, $t=3.187$ $p=.004$).

7.3.8. Retrospective memory (RM) for performing the PM task.

RM was higher in the younger group ($M=.85$, $SD=.15$) compared to the older group ($M=.55$, $SD=0.20$). A one-way analysis of variance (ANOVA) showed a statistically significant effect of group on the mean proportion of trials where participants' memory for the past PM performance was accurate [$F(1,49) = 32.904$, $p<.001$]. RM for past performance, was strongly influenced by whether or not the PM task was performed. Participants in both groups where more inclined to remember the trials where they had performed the PM task (younger: $M=.89$, $SD=.14$; older: $M = .63$, $SD = .20$) than the trials where they omitted to perform the PM task (younger: $M = .55$, $SD=.42$; older: $M = .27$, $SD = .26$).

A 2 by 2 mixed analysis of variance (ANOVA) with group as between-subjects variable and PM task (correct, or not) as within-subjects variable showed a statistically significant effect of PM performance [$F(1,42) = 44.077$, $p<.001$], a statistically significant effect of group [$F(1,42) = 17.375$, $p<.001$] but no significant effect of performance by group interaction.
The enhanced condition did not improve, in the long run, RM for past performance in either of the two groups. The mean proportion of trials where participants remembered their performance correctly was (for both groups) significantly higher in the enhanced condition (younger: $M=.92$, $SD=.15$; older $M=.67$, $SD=.29$) than in the standard condition (younger: $M=.79$, $SD=.24$; older: $M=.44$, $SD=.26$). But this difference is due to the fact that in the current experiment, the enhanced condition occurred, only after the PM task was performed. Therefore, it reflects the very strong effect of the PM performance on RM rather than the effect of the enhanced condition. When filtering out trials where PM task was omitted, the mean number of trials where participants remembered their performance correctly was very similar across conditions for both groups.

7.3.9. Confidence judgements on remembering PM performance (in trial questionnaire).

Results on the R/K encoding (confidence judgement questionnaire) showed that both younger and older participants reported to be very confident to remember that they had performed the PM task (very sure: 60% of the trials, 51% in the older group and 70% in the young group; sure: 30% of the trials, 55% in the older group and 40% in the young group).

The R/K encoding judgements performed during the trials were compared to the judgements (how certain they were that they had performed the PM task) participants gave at the end. The level of confidence was higher in the post experiment questionnaire, responses based on the highest confidence judgement score (recollection) accounting for 70% and 22% based on familiarity. Pearson correlation was significant in the young group only (young: r= .399, p<.001; correlation is significant at the 0.01 level (2-tailed)). Confidence ratings were similar across trials and for the two groups (young: M=4.51, SD=.72; old: M= 4.27, SD=.70).

7.3.10. RM failures.

Participants in both groups were more likely to misjudge their unsuccessful performance (trials where they omitted to perform the PM task) than their successful ones. The mean proportion of RM failures where participants forgot they have performed the PM task was M=.31, SD=.20 for the older group and M=.095, SD=12 for the younger. Therefore, younger participants misjudged 10 % of their successful performance while older participants misjudged 33%. The mean proportion of RM failures when participants forgot that they had omitted to perform the PM task was M=.69 (SD=.35) in the older group and M=.54 (SD.38) in the younger group. A 2 by 2 mixed analysis of variance (ANOVA) with group as between-subjects variable and type of RM failure as within-subjects variable showed a statistically significant effect of type of RM failure [F (1,42) = 56.460, p<.001], a statistically significant effect of group [F (1,42) = 7.693, p=.008] and no statistically significant group by type of memory failure interaction. Paired Sample t-test showed that the difference between the two types of RM failures is statistically significant in both groups: older [t (22) = -5.419, p<.001); younger (t (20) = -5.207, p<.001].

Younger participants were able to remember correctly twice as many of the trials where they have omitted to perform the PM task (M=.55, SD=.42) than the old group (M =.27, SD=.26). A one-way analysis of variance (ANOVA) showed that the main effect of group was statistically significant [F (1,43) = 7.263, p=.01].

Participants in both groups where able to remember the total number of OEs committed, significantly more accurately, than on which trials they have committed OEs (younger group: t (22) =-2.859, p=.009; older group: t (26) =-3.500, p=.002).

7.3.11. Confidence judgements on remembering PM performance (post trial questionnaire).

In the current study, both younger and older participants were more prone to make RM judgements based on recollection than on familiarity. The mean proportion of RM judgements based on recollection was significantly higher in the younger group ($M=.86$, $SD=.22$) compared to the older group ($M=.56$, $SD=.34$) [$t\ (45) = 3.753$, $p<.001$] while the mean proportion RM judgements based on familiarity was significantly higher in the older group ($M=.33$, $SD=.28$) compared to the younger group ($M=.09$, $SD=.15$) [$t\ (41) =-3.723$, $p<.001$]. In both groups, when the RM judgment was based on recollection, the mean number of correct RM responses was significantly higher that the incorrect responses but when based on familiarity, the difference between correct and incorrect responses was only significant in the younger group.

In order to assess the differential contribution of recollection and familiarity processes to the accuracy of the RM judgements, the mean proportion of judgements was measured based on recollection or familiarity for each of the correct and incorrect RM judgment (see Table 7.3). For both groups, recollection contributed more than familiarity to correct RM, but the difference was only statistically significant in the younger group [$t\ (22) =10.247$, $p<.001$]. RM failures were also based significantly more on recollection than familiarity in both groups (younger $t\ (14) =5.281$, $p<.001$, older: $t\ (25) =2.138$, $p=.042$). Both groups were prone to be very confident in their RM judgment even when incorrect.

RM response	Underlying memory process	Younger		Older	
Correct RM		Mean (N=23)	Std. Deviation	Mean (N=27)	Std. Deviation
	Correct recollection	.89	.18	.61	.33
	Correct familiarity	.11	.18	.38	.33
Incorrect RM		Mean (N=15)	Std. Deviation	Mean (N=26)	Std. Deviation
	Incorrect recollection	.73	.34	.50	.36

	Incorrect familiarity	.09	.17	.26	.24
	Don't know	.17	0.28	.26	.22

Table 7.3. The mean proportion and standard deviation of correct retrospective and incorrect retrospective memory judgements as a function of the underlying memory process (familiarity, recollection) and as a function of age.

Contrary to the initial prediction, the manipulation in the trials with the enhanced condition did not significantly increase confidence regarding RM for past performance in either group. The mean proportion of RM judgements based on recollection were similar across conditions in both groups (younger: standard condition: M=.86, SD=.24; enhanced condition: M=.88, SD=.23; older: standard condition: M=.58, SD=.34; enhanced condition M=.55, SD=.38). A linear model of RM confidence level as a function of condition was constructed; the model was not significant (F=.776, p=.379).

7.3.12. Results on the Prospective Retrospective Memory Questionnaire (PRMQ).

Independent samples t-test showed that the mean scores of younger and old participants did not differ significantly on either of the 3 scales of the questionnaire: RM scale (t (48) =104, p<.917); PM scale [t (48) =.588, p<.599]; total scale [t (48) =.371, p<.712]. Pearson correlations between age and scores on total (r=-.054), prospective (r=-.081) and retrospective scale (r=-.019) did not achieve statistical significance. There were no statistically significant correlations between scores on the PRMQ and participants' performance in the experiment for any of the variables (total number of trials where the PM task was performed, total number of REs, total number of OEs, accuracy of the memory for the PM task performance), however, this is not an unusual finding within the literature.

7.4. Discussion.

In the general context of an aging population, substantial efforts have been allocated to identifying intervention and strategies to increase older adults' autonomy and independent living. Habitual PM is essential for independent living therefore devising strategies to improve this function in older adults is of outmost importance. Two important studies McDaniel et al.'s (2009) and Marsh el al.'s (2007) have proposed strategies to enhance encoding of a habitual PM task in order to improve older adults' habitual PM performance and they reported conflicting results.

Experiment 7 further investigated the potential effect of enhanced encoding by employing the laboratory habitual PM paradigm devised by McDaniel et al.'s (2009) with several changes. Firstly, external stimuli were used to enhance encoding of the PM task, by displaying a different visual cue each time participants successfully performed the PM task, whereas McDaniel used a self-initiated stimulus, requiring participants to perform the same motor action each time after the PM task. The novel finding from this experiment is that enhancing encoding during a habitual PM action, even through the use of an external cue, also significantly improved older participants performance by reducing REs. Furthermore, the manipulation in the enhanced condition attenuated the age-related increase in REs as the PM task became more habitual. Therefore, older participants REs became more frequent in the standard condition only. This result extends the important findings previously reported by McDaniels et. al (2009) that older adults' REs in habitual PM task are significantly reduced by performing a self-initiated task. It is likely that these manipulations, which enhance the encoding of the PM action, could reflect changes in memory traces. Specifically, these manipulations probably provide additional contextual details, and thus result in an enriched, and thus stronger memory record.

Further supporting McDaniel et al.'s (2009) report, when analysing the pattern of PM errors as the task became habitual, the data from Experiment 7 showed that REs tended to increase and OEs tended to decrease, just for the older group. In the current study OEs still outnumbered REs across all trial blocks, in both standard and enhanced encoding conditions. This is a rather unusual finding for the literature, which reports that (under certain conditions, high level of ongoing task difficulty or divided attention) as tasks become habitual, older participants tend to commit more REs than OEs (Einstein et al., 1998; McDaniels et al., 2009). Several methodological differences may account for this discrepancy in the results; however, it should be first mentioned that investigating OEs/REs patterns was beyond the scope of this study. A notable difference consists in the nature of the experimental designs. In the current experiment the enhanced condition was intermixed within trials for both groups, the effect of the manipulation in the enhanced condition preventing REs was levelled between trials within a trial block. In McDaniels et al.'s (2009) study participants gave ratings on their PM performance after each trial, and the OEs that were correctly acknowledged were not taken into consideration. From the two pilot tests performed for Experiment 7 it emerged that by placing the (confidence judgement) questionnaire immediately after each trial reinforced the PM task, and prompted participants to perform it in the following trial, resulting in virtually no OEs in either group. Therefore, the design of the experiment was adapted in order to eliminate these possible effects: participants gave confidence judgement scores, immediately after the PM action and then the ongoing task on that trial was resumed. Thus, it is possible that this manipulation, in the enhanced condition, significantly decreased REs but may have had no significant effect in decreasing OEs. Furthermore, memory for the past performance was assessed for each trial at the end of the experiment, but all the OEs were taken into account even if they were acknowledged.

Regarding memory for past performance, in Experiment 7 participants were significantly more prone to remember correctly their successful performance than their unsuccessful performance. Enhancing encoding during the PM action did not improve, in the long run, either the accuracy or the confidence in the RM judgements for past performance for either group. This experiment identified a significant age-related decline in memory for past performance for both successful and unsuccessful PM actions. This age gap was significantly more pronounced for OEs: younger participants remembered 55% of the trials where they had omitted to perform the PM task; whereas older participants remembered only 27%. Additionally, the data show that not only did older participants commit more OEs but they also remembered fewer of the OEs they committed.

Similar results have been previously reported by Marsh et al. (2007). In a second experiment where they employed a manipulation meant to increase memorability of the PM response, older participants' memory for their past performance was less accurate than younger participants for both OEs and REs, with RM failures for OEs outnumbering RM failures for correct PM. The authors proposed that the manipulation in the enhanced condition may have created strong retrieval competition which exacerbated older participants' binding deficit: they have either failed to associate the PM action with the item during encoding or have forgotten the action-item association. Several findings in Experiment 7 also point to the fact that impairments in RM for past performance can be (at least partially) attributed to older participants' binding deficits. The analysis of participants' memory for their past performance showed that they were significantly more accurate in estimating the overall number of their OEs, than attributing an OE to a specific trial. Also, the manipulation in the enhanced encoding condition did reduce REs, suggesting that participants had actually associated the PM action with the trial. Because in Experiment 7 the questionnaire assessing RM was performed at the end of the 12 trials, it is possible that the older participants could have forgotten the PM action-trial association. This finding is in line with reports by several other authors. Hashtroudi, Johnson and Chrosniak (1989) proposed that in habitual PM tasks older adults may experience source monitoring difficulties, they may remember having performed a task but not the precise spatial and temporal context. Additionally, Kausler, Lichty and Davis (1985) reported age differences in identifying the time blocks in which certain tasks were performed. Craik (1986) has proposed that forming a coherent memory representation requires that multiple streams of information are conceptually bound together. Studies on episodic memory have already demonstrated that spontaneous encoding in older adults is more stereotypical. Older adults are prone to engage in less elaborate processes during encoding (Rabinowitz, Craik, & Ackerman, 1982, Dodson & Schacter, 2002) which negatively impacts their ability to build distinctive memory representations and to later discriminate between perceptually similar ones (Toner, Pirogovsky, Kirwan, & Gilbert, 2009; Stark, Yassa, Lacy, & Stark, 2013). It is possible that the manipulation in Experiment 7 was not strong enough to stimulate participants to effectively bind the information from the PM action episode together (the trial to the performance cue), even though the intention was to make each trial and each PM action cue as distinctive as possible. It was expected that a manipulation promoting a deeper (semantic as well as visual) association between trial and cue might facilitate higher quality processing and create a stronger bond between the trial and PM action, thus enabling a long-term effect of the manipulation.

An important factor in any memory test lies in the participants' ability to accurately evaluate their memory efficacy and predict their performance. This ability influences the efforts and individual strategies in which each participant engages, for encoding and remembering (Gilewski, Zelinski, & Schaie, 1990, cited in Mantyla, 2003). With this in mind, one of the other aims was to uncover the underlying mechanisms that drive these differences in performance. It was predicted that when participants made future oriented judgements regarding (just) performed PM actions their confidence to remember them will depend on the type of information that they encode along with the action. Specifically, participants will be more confident and accurate in their retrospective confidence judgements when the encoding of the PM action is accompanied by enriched contextual details (enhanced encoding condition) compared to their judgements when this information is not available (standard condition). Experiment 7 showed that for the enhanced condition trials participants were very confident they would remember having performed the PM task on a specific trial. Enhancing encoding during the PM action did not improve, in the long run, either the accuracy or the confidence in the RM judgements for past performance for either group. Younger participants were significantly more confident and accurate than older participants predicting their RM performance as well as in their RM judgements regarding their past performance.

A significant body of studies have found that in certain types of tasks, older participants overestimated their long-term memory performance (e.g. Murphy et al., 1981) as well as their short-term memory span (Bunnell, Baken, & Richards-Ward, 1999) and have suggested that overconfidence in PM performance abilities has a negative impact on the actual PM performance (Brandimonte, 2005). Two sources of evidence could support this 'overconfidence account' in the current sample. One deriving from the participants confidence judgements scores, where both groups tended to overestimate their performance on both PM and RM tasks. The other based on the contrast between self-reported memory performance to actual laboratory-based tests. The results of Experiment 7 showed that self-reported prospective and RM scores do not reflect the age-related differences identified with the laboratory experiment. Even though, based on the experiment performance, younger participants outperformed older participants in both PM and RM tasks, the self-rated PM and RM scores were similar for both groups. In this sample, older adults tended to highly overestimate their performance compared to younger adults. They were twice more prone to respond that they had performed the PM task when in fact they omitted it, and also performed 4 times more REs compared to the younger participants. It could therefore be possible that by being overconfident in their memory efficacy, older participants did not actively engage in encoding strategies even when environmental cues where available and when indirectly prompted towards encoding them.

There are some limitations in this study. With respect to assessing participants recollective and familiarity judgements of PM performance, it was only possible to analyse these for the trials with the enhanced encoding condition, as these were the only trials that had the R/K questionnaire. If an R/K question were to be included, even in the normal encoding trials, this could have potentially acted as an enhanced condition in itself. This issue was addressed by including another R/K questionnaire at the very end of the experiment in order to be able to assess what happened during those (normal encoding) trials. This last questionnaire also enabled the comparison of participants PM judgements with their RM judgements.

Another possible limitation is using a scaled confidence judgement rather than the classical R/K judgement. The results in the literature with regard to this matter are equivocal, as has been argued above, in the sample for Experiment 7 it seemed that the R/K method of questioning was not fully comprehended by participants, which lead to the choice of this alternative way of testing.

Lastly, the older participants in this study were unfamiliar with computer-based testing and a number of participants made comments in this regard (with some participants even dropping out for this reason). There is also a positive aspect to this, as it could be argued that this sample is more representative for this age group, and cultural background, than individuals that regularly take part in experimental research and almost become ‘professional test takers’.

CHAPTER 8: General Discussion.

8.1. Summary of main findings.

The outcome of each experiment reported in the thesis has been discussed individually in the relevant chapter (Chapters 2 to 7). Here I will discuss the general findings and their cumulative implications.

The main findings from the experiments reported in this thesis were:

(i) Repeated partial testing enhances long-term memory performance through a priming mechanism which can operate both at a story and at a sentence level.

(ii) Accelerated long-term forgetting is found associated with increasing age but not further determined by Alzheimer's disease.

(iii) Varying degrees of learning (DOL) only influences long-term memory performance when individual differences in learning performance are taken into account.

(iv) Long-term memory performance of prospective memory (PM) tasks (i.e., spontaneous retrieval) can be improved through repetition during encoding (e.g., additional learning). This result is similar in retrospective memory (RM) tasks (e.g., additional learning enhances RM performance), however, PM and RM task performance for younger adults do not correlate.

(v) The benefits (i.e., reducing repetition errors) of enhanced encoding appears in habitual PM tasks which are based on underlying recollection processes.

Each of these findings will be discussed in turn.

i. Repeated partial testing a priming effect.

Several studies have shown that even the testing of subparts of initially encoded material, not just the retrieval of the entire material, can have a positive effect on long-term memory performance (e.g., Chan, McDermott, & Roediger, 2006; Baddeley, Rawlings and Hayes, 2014; Baddeley, Allen, Atkinson & Kemp, 2019). When evaluating this effect, some authors have suggested that the benefits arising as a result of partial testing apply only to material that can be integrated, or reconstructed by participants (e.g., prose, video as opposed to individual words or pictures). With some exception (e.g., Rawlings and Hayes, 2014), few have assessed this effect in older and clinical populations, and fewer still have specifically devised studies to investigate what the cognitive basis of this enhancement effect is.

Chan, McDermott and Roediger (2006) proposed that this retrieval enhancement occurs due to an active search of related information, performed by participants during test. Chan (2009) further proposed that either the retrieval induced facilitation effect shares the same underlying mechanism (i.e., a conscious and active retrieval process) as the testing effect or it results due to a controlled process (i.e., spreading activation). Chan (2009) concludes that, even when information is retrieved in a collateral fashion (as in retrieval of sub-parts of material), it is strengthened over a long period of time (unlike semantic priming). The results from Experiments 1, 2 and 3 in this thesis, seem to suggest that this partial retrieval enhancement arises specifically as a result of implicit priming of existing representations.

As discussed in Chapters 2 and 3, both older individuals and individuals diagnosed with Alzheimer disease (AD) are known to have impaired learning but preserved priming (e.g., Camus et al., 2003; Lustig & Buckner, 2004; Bennett et al., 2006; Yano et al., 2008) that is why in Experiments 1, 2 and 3 older participants and people with AD were recruited to examine this priming hypothesis. In Experiment 1 and 2 it was hypothesised that if repeated partial testing enhancements arise because each test offers a new learning opportunity then older adults would benefit from it in a similar way as younger adults. This enhancement should also favour amnesic patients, if priming is key. Experiment 3 examined whether AD patients will benefit to the same extent as healthy controls (HC) from repeated partial testing, thus eliminating the difference in forgetting slopes between the two groups. Additionally, if repeated partial testing provides a new learning opportunity, individuals with learning deficits could potentially be mistaken as exhibiting accelerated long-term forgetting (ALF) since they would benefit from relearning to a lesser extent. This pattern should arise both when comparing younger to older adults as well as when comparing AD patients to HC. On the other hand, if repeated partial testing represents priming, then individuals with a lower learning capacity, such as older adults and AD patients, should exhibit relatively preserved long-term memory performance under repeated partial testing, as the act of repetition would serve to strengthen existing representations thus also benefiting them.

All three experiments (Experiment 1, 2 & 3) compared a condition where participants were tested repeatedly (3 delays over the course of 1 month) to a condition where participants were only tested after encoding and only once more at 1 month. This method allowed for a direct analysis of the effect of repeated partial testing on long-term memory performance, by separating the effect of testing from forgetting. Additionally, because of the previously mentioned view, that this facilitation effect is dependent in the level of material integration/coherence, Experiment 2, used material with a disrupted narrative (just in older and younger adults, not AD patients) to assess if the priming advantage remains. The aim was to assess if partial testing of material with a lower level of integration would benefit performance of younger and older adults to the same extent, thus suggesting that priming continues to operate at the sentence level, where different features of a given sentence are probed after different delays.

In Experiment 1 and 2, older participants took longer to learn the material compared to the younger groups. Similarly, in Experiment 3 AD patients took longer to learn the material compared to HC. Hence, in both cases the expected decline in learning ability was shown. Interestingly, the findings from all three experiments showed that participants were better able to retain information over the course of one month when tested repeatedly, as opposed to the condition when performance was only assessed immediately after learning and after 1-month only. Though in both conditions, all the groups declined in recall performance at 1-month test compared to post-encoding retrieval test, the decline was significantly reduced for all the groups in the condition with retrieval practice (this result holds for both younger, older and AD participants). This suggests that repeated-testing counters forgetting at 1-month delay, even when retesting does not involve relearning of the tested material, as different features of the initially encoded material were probed at each trial. The fact that AD patients benefit from repeated testing to the same extent as HC and older people as younger adults shows that, despite learning deficits, this sampling method delays forgetting. In turn, this may suggest that repeatedly retrieving sub-parts of material delays forgetting as a result of priming rather than relearning.

Compared to the majority of studies on practice effects, which used within subjects designs, all three experiments discussed in this section, employed a between subjects design which allowed for a separation of retesting effects from the effects of delay. This enabled a more accurate quantification of the magnitude of this effect by showing that performance is improved under repeated testing conditions, even with partial testing (sampling different features from each fable on every test session/delay) of both material with higher and lower levels of integration (narrative and sentence material).

Previous studies have indirectly implied that repeated testing might prove beneficial to AD patients. These studies reported that increasing the delays between various testing when recalling information repeatedly (spaced retrieval) improves memory performance in dementia patients and amnesiacs (e.g., Cull et al., 1996; Brush & Camp, 1998). To the best of my knowledge, Experiment 3 presents the first assessment of long-term forgetting in AD patients over an interval of one month. It is also the first study to compare forgetting rates in AD under a condition with retrieval practice to one without retrieval practice, showing that performance is improved under repeated testing conditions. Moreover, while other experiments aiming at studying retrieval practice in dementia patients have generally focused on simple cognitive tasks such as face-name associations, object-name or object-location associations, and cue-behaviour associations (for further examples see Chapter 3 and Creighton et al., 2013), the current study experiment used a complex task, which assessed associations between multiple features within stories.

ii. Accelerated long-term forgetting is found with increasing age but not in AD.

The methods used in Experiment 1, 2 and 3, employing a condition with multiple assessments, also allowed for an evaluation of forgetting curves across groups. Some authors have proposed that, in order to derive an accurate measure of forgetting, multiple retention tests, separated in time, are needed (e.g., Slamecka & McElree, 1983). Secondly, with regard to the choice of material, Elliot, Isaac and Muhlert (2014) recommend that when assessing accelerated long-term forgetting (ALF), repeated testing should be allowed while avoiding repeated retrieval of the same material as much as possible. Additionally, material should not demand excessive (initial) learning time and should allow for different subsamples of questions to be tested via cued recall after a range of delays (Elliot, Isaac & Muhlert, 2014). Experiments 1, 2 and 3 used material that was specifically designed to comply with these recommendations. By using material that minimises relearning by sampling different features on every subsequent test/delay (with no feedback), which also avoids ceiling or floor effect (in healthy participants) and comparing a condition with retrieval practice (testing on: immediate, 24 hours, one week and one month) to a condition without retrieval practice (only immediate and 1-month tests), the assessment of ALF was facilitated. The results from these experiments showed that, while AD patients had equivalent forgetting rates to HC, older individuals showed accelerated forgetting compared to younger adults (observed in the condition without retrieval practice).

The results from Experiment 3 speak against the occurrence of ALF in AD patients, when comparing performance from post-encoding retrieval to 1-month test, AD patients did not show ALF in either the condition with retrieval practice or the condition without retrieval practice. AD patients in Experiment 3 did differ from HC in learning ability, as they needed more trials to reach criterion at encoding compared to HC. Because differences in initial learning ability have been proposed to confound analyses of forgetting rates (Loftus, 1985) and/or may lead to underestimations of forgetting in lower-performing groups, all three experiments (Experiment 1, 2 & 3) aimed to train all participants to a pre-set criterion (70% correct). All participants reached this criterion (after varying encoding trials), apart from seven AD patients who did not, and were excluded from the statistical analysis. The finding that AD is not characterised by ALF, but rather by an encoding impairment, has been previously reported in the literature. The difference between the studies reporting ALF and those which do not has been proposed to be due to a failure to equate baseline performance between groups. There have also been reports from studies which did equate performance (or did not need to e.g., Weston et al., 2018) yet still find ALF for AD patients.

A recent study by Weston et al. (2018) investigating a group of participants diagnosed with autosomal dominant (familial) AD, due to an inherited gene mutation, found that ALF may be an early pre-symptomatic feature predating other amnesic deficits. Though the participants in their study had equivalent learning ability and recall performance over a short interval when reassessed after a longer delay (7 days) pre-symptomatic mutation carriers had forgotten more than had non-carriers (Weston et al., 2018). A possible explanation for the difference between their results and those from Experiment 3 (Chapter 3) may lie in the choice of groups. Specifically, it may be that forgetting patterns are the same when we compare AD patients to older adults, however, if we were to include a young-old group (i.e., 40-50 years old) we may find a difference in forgetting rates. We know that the pattern and amount of forgetting in older populations differs with the type of memory being tested. Several studies have additionally proposed that ‘older adults’ does not represent a homogeneous group in terms of memory performance and that significant differences in forgetting rates can be found between subgroups of older adults when employing narrower age ranges.

Kliegel and Jäger (2006) compared memory performance on a PM task between four age groups, a young-young (22–31 years), a young-old (60–69 years), a middle-old (70–79 years) and an old-old (80–91 years) and found that only the young-old age group differed significantly from the younger group. While, to the best of my knowledge, no similar patterns of performance have been reported in the literature investigating RM, it may be that Kliegel and Jäger’s (2006) finding could explain the results from the experiments discussed in the current section. Specifically, the older adults in Experiments 1 and 2 would fall in the category of young-old (60–69 years) while the healthy controls (older adults) from Experiment 3 would fall in the category of old-old (80–91 years), thus potentially explaining the finding of ALF in the healthy old sample but not in AD.

iii. Degrees of learning (DOL) and subsequent forgetting.

Forgetting rates may be misinterpreted when comparing groups that are performing at different levels, if baseline performance is not equated. Though this issue has been frequently raised in the forgetting literature, other influential studies have shown that the degree of initial learning has no effect on subsequent forgetting rates. As is the case with many experimental studies, this difference in findings may derive from differences in the method and analysis employed to evaluate the topic under investigation. This choice (of method and analysis employed) usually stems from authors’ views on the theoretical and methodological perspectives of the topic under investigation.

For example, Slamecka and McElree (1983) and Loftus (1985a/b) used different methods and definitions for assessing the effect of DOL on forgetting, due to their differing views on the theoretical and methodological perspectives on the relation between the amount of original learning and forgetting. Other examples of different theories and methods of analysis were subsequently proposed by Bogartz (1990) and Yang et al. (2016). Bogartz (1990) used a different model of forgetting (i.e., a psychological function), yet his findings still supported those reported by Slamecka and McElree (1983). Yang et al. (2016) used the same analysis as Slamecka and McElree (1983) and found the opposite result. The comparison of time intervals in Yang et al.'s (2016) study was within participants, while in Slamecka and McElree's (1983) study, it was between participants, which may have led to this difference in findings.

Two experiments were designed to investigate the issue of DOL in the current thesis, Experiment 4 and 5, which will be discussed in turn. Experiment 4 successfully replicated Slamecka and McElree's (1983) original study (Experiment 1 of 3) and found that DOL does not influence forgetting rate. The aim of Experiment 4 was not so much to settle the dispute between the opposing viewpoints with respect to DOL, but rather to verify the accuracy of the findings reported by Slamecka and McElree (1983). Therefore, the experiment employed exactly their same methods and procedures; the only deviation was in an additional retention test (after 10 days).

Replications are very much needed in the field of psychology in order to verify the reliability of the original results. Of course, choosing a different analytical and methodological strategy to investigate the same topic can be useful in explaining new theoretical assumptions. The next experiment was therefore designed to investigate a different view on how DOL affects forgetting.

The choice of theory and methods employed in Experiment 5 derived from the idea that differences in retention might derive from individual differences in learning, specifically that faster learners may forget less than slower learners (Gillette, 1936; McGeoch & Irion, 1952). Thus, Experiment 5 hypothesised that not only the number of trials at encoding (varying DOL) but also each individual's learning capacity will influence subsequent forgetting rates. Additionally, because within group variability may influence the interpretation of group forgetting rates (MacDonald et al., 2006), the effect of DOL was investigated in groups that were chosen to be very similar in terms of learning capacity (for a more detailed review of the methodology see Experiment 5). Results from this experiment suggested that faster learners (i.e., participants exposed to higher DOL) have slower rates of forgetting compared to those with lower DOL. Instead, the rate of forgetting is similar for slower learners when exposed to higher or lower DOL. Thus, the results from Experiment 5 partly replicate those from Experiment 4 and from previous findings (e.g., Slamecka & McElree's 1983). A possible explanation as to why slower and faster learners are affected differently can be drawn from looking at how participants retain individual items. The pattern of results after running an item (feature-association) analysis would suggest that the main difference between slower and faster learners is driven by differences in the consolidation of items encountered during the second trial. Specifically, faster learners retain items that are encountered on a second trial at encoding (higher DOL) over the course of one month. While the slower learners forget these items and only retain the ones gained on the first encoding trial, in a sense slower learners do not benefit from higher DOL over the long-term. Gentile et al. (1995) propose that a plausible prediction is that faster learners will show a memory advantage when afforded the opportunity to reorganise previously encountered material, as they are better at adopting organisation strategies.

iv. Long-term memory performance of prospective memory (PM) tasks is also improved through repetition during encoding.

PM failures lead to frequent complaints in everyday life, thus a lot of research has been dedicated to understanding the sources of these failures and finding strategies to reduce them. Experiments 6 and 7 investigated how enhanced encoding strategies may mitigate PM performance in an event-based PM task and in a habitual PM task. PM has been classified as a distinct form of memory, though the term PM is an 'umbrella term' covering different types of PM memory. As such, different types of PM may lead to different types of PM failures. In episodic PM, the difficulty consists in self-initiation of an action in response to a target cue (event/time). In habitual PM, the difficulty, especially for older individuals, lies in remembering that the action has already been performed and not repeating it. These kinds of tasks are at the opposing end of memory hierarchy theory (Craik, 1986); one may employ self-initiated processes while the other relies heavily on contextual factors in the environment. As such, Experiments 6 and 7 were devised with the premise that interventions meant to enhance PM performance should address specific difficulties associated with each type of PM failure.

Experiment 6 compared PM performance at 1-month on an event-based PM task between younger participants who encoded a PM intention under a standard condition and under an enhanced encoding condition (repeated the PM intention 3 times). Memory performance at 1-month was assessed separately for the retrospective and the prospective components of the PM task. Research has only recently focused on investigating the efficacy of strategies meant to enhance encoding during the intention formation stage of a PM task (Chasteen, Park, & Schwarz, 2001; Zimmermann & Meier, 2010; Altgassen et al., 2016). For example, implementation intention (IIE) is an encoding strategy frequently used in goal-oriented behaviour which provided mixed results in the context of PM. A recent study by Altgassen, Kretschmer and Schnitzspahn (2016) compared IIE to a condition where participants repeated the PM instruction during intention formation phase and showed that the latter condition was more efficient in increasing PM performance in younger adults compared to IIE strategy. They employed a laboratory experiment which was conducted only over a relatively short delay between intention formation and retrieval of the PM task.

Experiment 6 investigated whether enhancing encoding during the intention-formation phase, by asking the participants to repeat the intention during encoding, enhances performance over a long interval. The results showed improved overall PM performance at one month, compared to a standard encoding condition. Thus, repeated encoding of the intention proved to be an efficient strategy for enhancing PM performance over long retention intervals. A second finding of this experiment was that, even when a PM task is performed over long delays, the effect of the enhanced encoding manipulation derives for strengthening the prospective and not the retrospective component of the PM task. Similar findings were reported by Zimmermann and Meier (2010) who employed IIE as an encoding strategy using a laboratory paradigm over a short delay.

The analysis of the prospective and retrospective components also revealed that the main failure in PM tasks derives from the self-initiation (or lack of) the PM response/action even when the association between the target event and the action to be performed is correctly encoded. Irrespective of the encoding condition, 43 out of 46 participants (97%) successfully stored and retained the association between the prospective cue and the intended action (remembered what they were supposed to do and when). Only 25 of them were, however, able to successfully self-initiate the PM response (to deliver a message without a prompt from the experimenter). Studies performed in laboratory settings (Einstein & McDaniel, 1990) have long established that the retrospective component of a PM task is significantly easier to remember than the PM component (Dismukes, 2010). Experiment 6 showed that even at long retention intervals (1 month) the relative contribution of retrospective and prospective components to PM does not reverse, with the self-initiated PM component being the main contributor to the PM failures. It should be mentioned that the retrospective memory component of the PM task in Experiment 6 was relatively easy, it only included one salient target event and the action to be performed was also easy. Several laboratory studies have shown that the contribution of retrospective memory component increases with the difficulty of the target cue and the intention (Einstein & McDaniel, 1990; Smith et al., 2014).

The underlying mechanisms that account for the effect of different strategies to enhance encoding of the intention on future PM performance are still debated. In the case of an IIE strategy, several laboratory studies have proposed that it stimulates spontaneous retrieval of the PM task (e.g. Einstein & McDaniel, 2010; Rummel, Einstein & Rampey, 2012) by increasing the perceived importance of the task (e.g., Brewer & Marsh, 2010; Smith et al., 2014) or by strengthening the association between the intended action and the target event/cue (e.g. Gollwitzer 1999, McDaniel et al., 2008; McDaniel & Scullin, 2010). Taking into consideration the results from Experiment 6, it is unlikely that the positive effect of the encoding strategy was due to it enhancing retrospective recall of the intention, as this was similar in both conditions. All but 2 participants were able to recall the PM intention (when and what they were supported to do) after being given a very neutral prompt ('Was there something else'). All participants in both encoding conditions recalled the retrospective component while only half of them were able to self-initiate the PM response. This suggests that the retrospective component is neither the source of the failure in PM tasks nor the component where the facilitating effect of enhanced encoding on future PM performances originates.

The most likely explanation for the results of Experiment 6 is that repeated encoding of the intention facilitated self-initiated retrieval of the PM response by increasing the specificity of the PM event (receiving a call from the experimenter) as a cue singling that the intended action has to be performed (tell the experimenter that: 'It is a sunny day'). In other words, the encoding manipulation may have reinforced the perception that the main purpose of the call from the experimenter was to initiate the PM response from the participant, thus facilitating self-initiated retrieval of the PM response.

Lastly, Experiment 6 showed that performance on the PM task was not correlated with performance on a RM task (for details of the RM task see Experiment 6). These findings are in line with previous research consistently reporting no correlation between RM and PM (e.g., Wilkens & Baddeley, 1978; Einstein & McDaniel, 1990; Shelton et al., 2016). The existence, or lack of, a correlation between PM and RM performance has been at the heart of the debate on PM being a distinct form of memory. Graf (2001) proposes two attributes that limit the conclusive assessment in distinguishing between PM and RM: the fact that existing methods yield binary data, and that PM performance is not ‘process pure’ (p. 539), namely a PM tasks includes both prospective and retrospective memory components. Experiment 6 managed to overcome one these problems. As all participants remembered the retrospective component of the PM task, there was no retrospective memory effect on PM performance, therefore it could be claimed as ‘pure’ PM.

v. The benefits of enhanced encoding in habitual PM tasks may be based on underlying recollection processes.

Experiment 7 employed a laboratory paradigm (adapted from McDaniels et al., 2009) that was designed to compare performance on a habitual PM task between younger and older adults in an enhanced condition and a standard condition. In the enhanced condition, after participants performed the PM task an external cue (the sentence ‘You have clicked’ and a coloured picture - a different one for each trial) was displayed on the computer screen. This manipulation significantly improved older participants performance by reducing repetition errors (REs). Furthermore, the manipulation in the enhanced condition attenuated the age-related increase in REs as the PM task became more habitual. Therefore, older participants REs became more frequent in the standard condition only. Only two other studies investigated strategies to enhance PM performance during retrieval (McDaniels et al., 2009; Marsh et al., 2007). Both studies employed self-initiated motor actions as the retrieval enhancement strategy and reported conflicting results. The results from Experiment 7 are in line with findings reported by McDaniels et al. (2009) who found that REs in habitual PM significantly decreased when performance of the PM task was accompanied by a self-initiated motor action (instructing participants to raise their arm above their head after performing the PM tasks).

PM failures in habitual tasks have been attributed to age related difficulties in internal-source monitoring (confusing the action with the thought or intention of performing a certain task: e.g., Maylor, 1996; Johnson, Raye, & Estes, 1981) and temporal discrimination (confusing a recently performed action with a previous one: e.g., Friedman, 1993). Time-based PM task are self-initiated and provide very little environmental support compared to event-based tasks where external cues facilitate retrieval of the intended action. Building on this effect, it was hypothesised that external cues may be employed to support performance in a habitual PM task, in order to increase memorability of the act of performing the PM action (during the retrieval stage/its execution) thus decreasing REs. It was expected that by using a manipulation promoting a deeper association (semantic as well as visual) between trial and cue, higher quality processing may be facilitated creating a stronger bond between the trial and PM action, thus enabling a long-term effect of the manipulation.

McDaniels et al. (2009) proposed that the enhanced retrieval manipulation facilitated reduction of REs either by increasing attentional resources allocated to performing the PM task or by providing enhanced sensory information about the act of performing the PM task, thus resulting in a more distinct memory record. The manipulation to enhance the encoding of the PM action (used in Experiment 7) during retrieval was different from that of McDaniels, using an externally generated cue which allowed for the presentation of a different performance cue for each PM response. A study by Sadeh, Ozubko, Winocur and Moscovitch (2014) proposed that the way we forget may depend on how we remember, the characteristics of forgetting being influenced by the underlying declarative memory representations: recollection or familiarity (for a more detailed review see Chapter 7). Seen as in habitual tasks, the necessity of initiating (or not) a certain action is highly dependent on the accurate memory of the previously performed action (Marsh et al., 2007; McDaniel et al., 2009). In Experiment 7 it was hypothesised that the manipulation will most likely decrease REs by providing additional contextual details, and thus resulting in an enriched, stronger memory record for the performed PM action in each particular trial, presumably by 'making' participants remember the PM action through a recollection process rather than a familiarity one. It's been proposed that true recognition of a previously performed action/task requires memory for detailed contextual information (presumably such as in the case of processes supported by recollection), while false recognition is based on a feeling of "déjà vu" (Brainerd & Reyna, 2002).

The manipulation in Experiment 7 was successful in enhancing short-term memory for past performance (as shown in the reduction of REs). It did not also enhance long-term memory for past performance, which was assessed at the end of the experiment for each of the 12 trials. Additionally, the older adults in this sample overestimated both their future PM performance and their ability to remember their past PM performance. The trial by trial assessment of participants confidence in remembering that they have performed the PM task yielded very high confidence levels. The post-experiment assessment revealed that, especially older participants, could not accurately recall their past performance during each trial and were more prone to remember that they had performed the PM task when in fact they had not. This comparison clearly shows a diminished ability to predict future memory performance, that may result from the fact that when planning to recall something in the future, people tend ‘to anchor on their current state’ (Meyvis, Levav & Ratner, 2010, p. 579) and be biased toward assuming that memory will remain stable over time (Koriat et al., 2004).

People’s ability to accurately evaluate their memory efficacy and predict their performance is important in the context of habitual PM as it influences the efforts and individual strategies they allocate to encoding and remembering (Gilewski, Zelinski, & Schaie, 1990, cited in Mantyla, 2003). Overconfidence in PM performance abilities has a negative impact on the actual PM performance (Brandimonte, 2005). Indeed, subjective memory performance with regard to both PM and RM components, as measured with the Prospective and Retrospective Memory Questionnaire (PRMQ; Smith, Della Sala, Logie, & Maylor, 2000), did not correlate with objective memory performance measured with the laboratory experiment. The age-related deficits in performance identified during the laboratory experiments were not found when comparing self-assessment scores. This is a usual finding reported in literature. The PRMQ investigated participants’ self-assessed PM performance in daily living. The usual finding in the literature is that older adults’ PM performance is poorer compared that of younger adults when assessed in laboratory experiments and better when assessed in more naturalistic settings. This pattern of findings is called ‘the age–PM-paradox’ (Rendell & Craik, 2000 cited in Schnitzspahn, Kvavilashvili & Altgassen, 2018). This term could suggest that older adults do not overestimate their PM capacity but rather outperform younger adults in daily life. The results from Experiment 7 suggest that older adults overestimate their performance based on the discrepancy between their evaluation regarding memory for the past performance (assessed through an in-trial questionnaire) and actual RM for the past performance (assessed at the end of the experiment). These results, however, cannot exclude the possibility that this discrepancy resulted from older adults underestimating the difficulty of the RM task.

The pattern of findings from Experiment 7 shows that creating a richer contextual environment during performance of a PM task can decrease REs. Therefore, teaching older adults to gather richer traces from the environment each time they perform a habitual PM task may prove to be an efficient method for decreasing REs in daily life also. This is particularly important, in light of Experiment 7 other finding, that older adults overestimate their RM and PM capacity.

8.2. Contributions of the experiments combined.

The experiments presented in this thesis showed that repeated partial testing enhances long-term memory performance for younger adults, older adults (Experiments 1 & 2) and people diagnosed with AD (Experiment 3) when tested both on material with a high level of integration (i.e., narrative material - Experiments 1 & 3) as well as with material with lower levels of integration (i.e., unrelated sentences - Experiment 2). The fact that repeated retrieval of subparts of material enhanced memory performance for both older adults and AD patients, despite their learning deficits, shows that this sampling method delays forgetting through a priming mechanism rather than a relearning mechanism. Additionally, the results from Experiment 2, which used material with a disrupted narrative (sentences), yet showed the same memory enhancement, suggest that the priming mechanism can operate at a sentence level not just at a story level. The design of Experiments 1-3 allowed for a comparison of forgetting rates between groups. Experiments 1 and 2 showed that older participants forget at a faster rate compared to younger adults, while Experiment 3 provides evidence against AD patients showing accelerated forgetting compared to healthy age matched controls. Though getting AD patients to reach the pre-set level of performance on the post-encoding retrieval test (learning criterion of 70%) required exposing them to significantly more learning trials, the subsequent rate of forgetting over the course of 1-month was similar to that of HC (in both conditions). This pattern of results suggests that once baseline performance is equated, AD patients do not show ALF.

Experiments 4 and 5 were designed to investigate the need, or lack of, equating baseline performance between groups performing at different levels. Though recent research emphasises the importance of equating performance when measuring forgetting rates (e.g., Elliot, Isaac & Muhlert, 2014), older studies concluded that equating initial performance between groups is not necessary, as the course of forgetting is independent of degree of initial learning (e.g., Slamecka & McElree, 1983). Therefore, Experiment 4 was designed to see whether Slamecka and McElree’s (1983) results would replicate. By employing the same methods and design (with an additional retention interval) as Slamecka and McElree's, Experiment 4 found that varying the DOL does not influence the forgetting rate. It was then hypothesised that if individual differences in learning capacity, as well as the number of trials at encoding (varying DOL) are taken into account, we may find that within group variability will influence subsequent forgetting rates. Thus, Experiment 5 assessed the forgetting rates of homogeneous groups (individuals which were chosen to be very similar in terms of learning capacity) after varying the DOL. When separate groups of faster and slower learners were exposed to different DOL (higher vs. lower) the results showed that subsequent forgetting rates are differentially affected. This was only the case of faster learners, as the rate of forgetting remained similar for slower learners both when exposed to higher or lower DOL. Additionally, Experiment 5 showed that slower learners only benefited in the short-term (higher performance on the second trial) but not in the long-term, when exposed to higher DOL. Specifically, in the high DOL condition, though slower learners reached a similar level of performance after the second encoding trial as that of the fast learners, performance at 1-month was still higher in the faster learning group. While the results of Experiments 4 and 5 cannot settle the debate on whether or not baseline performance should be equated, the results from Experiment 5 can at least prove useful in avoiding scaling problems, as they showed that groups can be taught to the same initial score while at the same time not changing the slower learners' long-term memory performance.

Finally, the last two experiments in this thesis looked at forgetting in a different memory system, specifically PM. Experiment 6 showed that higher degrees of learning of an intention (by repeating it) during its formation stage can enhance long-term memory performance for an event-based PM task. Additionally, the results showed that even when looking at PM over a long retention interval, the initiation of the PM action still remains the most challenging component of the task, and the main source of failures in PM. The separate analysis of memory performance for the retrospective and prospective memory components of the PM task showed that the retrospective component was similar irrespective of the encoding condition. This suggests that the enhancing strategy facilitated PM by strengthening the connotation of the target event as a cue to initiate the PM action. Finally, younger adults' PM memory performance is not correlated with RM performance (on a cued recall task), when both are assessed over the course of one month. Experiments 7 assessed the ability of younger and older adults to perform an intention in the context of routine everyday tasks, specifically in a habitual PM task. In habitual PM tasks, the necessity of initiating (or not) a certain action is highly dependent on the accurate memory of the previously performed action (Marsh et al., 2007; McDaniel et al., 2009). Thus the encoding phase of the PM tasks in Experiment 7 was manipulated in an attempt to reduce interference and promote the encoding of detailed contextual information. The results of this experiment showed that enhancing encoding during the execution of the PM habitual action, by using an external stimulus, can increase younger and older adults memory performance and thus reduce the likelihood of them committing REs.

8.3. How to measure forgetting and the analysis of forgetting data.

There are many approaches to evaluate data that bear on a research question (Gelman & Loken, 2016). This diversity in analytic choice may result in different conclusions when the same data are analysed by using different procedures (Silberzahn et al., 2018). Such a case is well illustrated in the previously mentioned debate between Loftus and Slamecka, whereby each author's definitions of how forgetting is affected by the degree of initial learning (DOL) led them to different analytic strategies and consequently to different conclusions.

Slamecka (1985; Slamecka & McElree, 1983) defined forgetting as a decrement in performance. Accordingly, the forgetting rate would be the rate of change in performance as a function of time. Slamecka used a fairly customary procedure for assessing whether groups differ in forgetting rates, by testing for significant interactions between a retention interval and a treatment variable (ANOVA) (Slamecka, 1985). The test establishes whether the slopes of the respective retention functions differ from one another (using vertical comparison of forgetting). The slopes represent the rate of forgetting associated with each level of the independent variable and if they do not reliably differ (e.g., if one slope does not become shallower) then forgetting is said to be equal between groups.

Loftus criticised this approach, proposing that the vertical comparison method favoured by Slamecka (1985) is subject to scaling problems that yield inconsistent conclusions both within and across experiments. He proposed to assess the effect of original learning (or any other variable) on forgetting as a horizontal interaction between variables. Loftus (1985) defined forgetting as a change from a state with greater retention to one with less retention, with performance being in direct relation with mental state. This implied horizontal parallelism as a measure of identity of forgetting rate.

Therefore, while Slamecka's approach focuses on differences between the dependent variable at a fixed level of the independent variable (vertical difference – comparing performance levels between groups at various delays), Loftus' s approach focuses on differences between the independent variable at the fixed level of the dependent variable (horizontal differences - comparing the length of the delay necessary for groups to reach the same level of performance). Loftus advocates that instead of assessing whether differences in performance between groups measured at a given delay will remain the same, we should assess whether the retention interval difference at which performance is equal between groups remains the same for all equal performance levels. He claims that if the latter will be constant, then there is horizontal parallelism; if it is not, there is no horizontal parallelism, and there is an interaction. Slamecka's (1985) rebuttal to Loftus's comments, emphasised that the horizontal comparison method requires that DOL be confounded with list age, to which Loftus replied that Slamecka's vertical comparison method requires that DOL be confounded with the performance level over which forgetting is assessed.

Despite Loftus's attempts to criticise the vertical comparison method, many statistical textbooks describe that an interaction is indicated when "the lines are not parallel" (vertical difference, as Slamecka considered). Thus, all psychologists are familiar with the concept of an interaction and often report and interpret interactions obtained in their own experiments. Wagenmakers and colleagues (2012), who also criticised this approach, have proposed to use matching procedures which would ensure that the forgetting curves (of different groups) overlap initially thus greatly increasing the opportunity to observe the true nature of an interaction.

The data in the present thesis were analysed taking into account all the points above. I used (where applicable) the 'traditional approach' of comparing for parallelism (as in Slamecka). However, I have also accounted for the problems that this method may raise (as pointed out by Loftus) by equating baseline performance. The different groups of participants in Experiments 1, 2 and 3 were matched on initial learning. The remaining experiments in my thesis were either specifically designed to control for possible scaling issues from the start (by splitting participants into groups based on similar initial scores -

Experiment 4); or investigated topics which are not assessed based on a significant interaction between the retention interval and the treatment variable.

REFERENCES.

Abbott, E. E., 1909. On the analysis of the factors of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1), 159.

Abrahams, J.P., Camp, C.J., 1993. Maintenance and generalization of object naming training in anomia associated with degenerative dementia. *Clinical Gerontology*, 12, 57–72.

Adams, H. E., Peacock, L. J., 1965. Retroactive amnesia from electroconvulsive shock: Consolidation disruption or interference?. *Psychonomic Science*, 3(1-12), 37-38.

Alber, J., 2014. Improving Longer-Term Memory via Wakeful Rest in Health and Amnesia: Evidence for Memory Consolidation. The University of Edinburgh, Edinburgh UK (Unpublished doctoral dissertation).

Altgassen, M., Kretschmer, A., Schnitzspahn, K. M., 2017. Future thinking instructions improve prospective memory performance in adolescents. *Child Neuropsychology*, 23(5), 536-553.

Altmann, E. M., Schunn, C. D., 2012. Decay versus interference: A new look at an old interaction. *Psychological Science*, 23(11), 1435-1437.

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, fourth ed. Author, Washington (DC). text rev.

Amlund, J. T., Kardash, C. A. M., Kulhavy, R. W., 1986. Repetitive reading and recall of expository text. *Reading Research Quarterly*, 21, 49-58.

Anderson, M. C., Bjork, R. A., 1994. Mechanisms of inhibition in long-term memory: A new taxonomy.

Anderson, M. C., Bjork, R. A., Bjork, E. L., 1994. Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063.

Anderson, M. C., Neely, J. H., 1996. Interference and inhibition in memory retrieval. In Memory (pp. 237-313). Academic Press.

Anderson, M. C., McCulloch, K. C., 1999. Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 608.

Anderson, M. C., Green, C., McCulloch, K. C., 2000. Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 26(5), 1141–1159.

Anderson, M. C., 2003. Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory & Language*. 49, 415-445.

Anderson, R. G., Greene, W. H., McCullough, B. D., Vinod, H. D., 2008. The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, 15(1), 99-119.

Andrés, P., Howard, C. E., 2011. Part set cuing in older adults: Further evidence of intact forgetting in aging. *Aging, Neuropsychology, and Cognition*, 18(4), 385-395.

Arnold, N. R., Bayen, U. J., 2019. Prospective memory: Comparing self-and proxy-reports with cognitive modelling of task performance. *Journal of Applied Research in Memory and Cognition*, 8(2), 244-254.

Baddeley, A. D., Warrington, E. K., 1970. Amnesia and the distinction between long- and short-term memory. *Journal of verbal learning and verbal behavior*, 9(2), 176-189.

Baddeley, A. D., Wilkins, A., 1984. Taking memory out of the laboratory. In J.E. Harris & P.E. Morris (Eds.), Everyday memory, actions and absent-mindedness, 1-17.

Baddeley, A.D., 1986. Working memory. Oxford: Oxford University Press.

Baddeley, A.D., Rawlings, B., Hayes, A., 2014. Constrained prose recall and the assessment of long-term forgetting: the case of ageing and the Crimes Test. *Memory*, 22 (8), 1–8.

Baddeley, A., Eysenck, M., Anderson, M., 2015. *Memory*. London: Psychology Press.

Baddeley, A.D., Atkinson, A., Kemp, S., Allen, R., 2019. The problem of detecting long- term forgetting: evidence from the crimes test and the four doors test. *Cortex*, 1–8.

Bailey, N. W., Thomson, R. H., Hoy, K. E., Hernandez-Pavon, J. C., Fitzgerald, P. B., 2016. TDCS increases cortical excitability: direct evidence from TMS-EEG. *Cortex*, 74, 320-322.

Balota, D. A., 1983. Automatic semantic activation and episodic memory encoding. *Journal of verbal learning and verbal behavior*, 22(1), 88-104.

Balota, D. A., Duchek, J. M., 1988. Age-related differences in lexical access, spreading activation, and simple pronunciation. *Psychology and Aging*, 3(1), 84.

Barker, W.W., Luis, C.A., Kashuba, A., Luis, M., Harwood, D.G., Loewenstein, D., et al., 2002. Relative frequencies of Alzheimer’s disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Disease and Associated Disorders*, 16, 203–12.

Barrouillet, P., Bernardin, S., Camos, V., 2004. Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83.

Bartlett, F. C., 1932. Experiments on remembering: The method of repeated reproduction. Remembering: A study in experimental and social psychology, *Cambridge. University Press*, 63-94.

Bateman, R. J., Aisen, P. S., De Strooper, B., Fox, N. C., Lemere, C. A., Ringman, J. M., ... & Xiong, C., 2011. Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease. *Alzheimer's research & therapy*, 3(1), 1.

Battig, W. F., Montague, W. E., 1969. Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of experimental Psychology*, 80(3p2), 1.

Bäuml, K.H., Hartinger, A., 2002. On the role of item similarity in retrieval-induced forgetting. Memory, 10(3), 215–224.

Bäuml, K. H. T., Samenieh, A., 2010. The two faces of memory retrieval. *Psychological Science*, 21(6), 793-795.

Bäuml, K. H. T., 2019. Context retrieval as a critical component in selective memory retrieval. *Current Directions in Psychological Science*, 28(2), 177–182.

Beck, S. M., Ruge, H., Walser, M., Goschke, T., 2014. The functional neuroanatomy of spontaneous retrieval and strategic monitoring of delayed intentions. *Neuropsychologia*, 52, 37-50.

Becker, J.T., Boller, F., Saxton, J., McGonigle-Gibson, K.L., 1987. Normal rates of forgetting of verbal and non-verbal material in Alzheimer’s disease. *Cortex*, 23 (1), 59–72.

Bell, B. D., Fine, J., Dow, C., Seidenberg, M., Hermann, B. P., 2005. Temporal lobe epilepsy and the selective reminding test: the conventional 30-minute delay suffices. *Psychological Assessment*, 17(1), 103.

Bell, B. D., Giovagnoli, A. R., 2007. Recent innovative studies of memory in temporal lobe epilepsy. *Neuropsychology review*, 17(4), 455-476.

Bennett, I.J., Golob, E.J., Parker, E.S., & Starr, A., 2006. Memory evaluation in mild cognitive impairment using recall and recognition tests. *Journal of Clinical and Experimental Neuropsychology*, 28, 1408-1422.

Bergman, E.T., Roediger, H.L., 1999. Can Bartlett’s repeated reproduction experiments be replicated? *Memory & Cognition*, 27: 937–947.

Birt, A. R., 2001. Prospective memory, a distinct form of remembering?: subtitle evidence from task comparisons and normal aging (Doctoral dissertation, University of British Columbia).

Bjork, R. A., 1994. Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185–205). Cambridge, MA: MIT Press.

Blackburn, H., Benton, A., Shaffer, L.F., 1957. Revised administration and scoring of the Digit span test. *Journal of Consulting and Clinical Psychology*, 21 (2), 139–143.

Blake, R. V., Wroe, S. J., Breen, E. K., McCarthy, R. A., 2000. Accelerated forgetting in patients with epilepsy: evidence for an impairment in memory consolidation. *Brain*, 123(3), 472-483.

Bogartz, R. S., 1990. Evaluating forgetting curves psychologically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 138.

Bouwmeester, S., Verkoeijen, P. P. J. L., 2011. Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language*, 65(1), 32–41.

Braak, H., Braak, E., Yilmazer, D., De Vos, R. A. I., Jansen, E. N. H., Bohl, J., 1996. Pattern of brain destruction in Parkinson's and Alzheimer's diseases. *Journal of neural transmission*, 103(4), 455-490.

Brainerd, C., Reyna, V., 2002. Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, 11(5), 164-169.

Brandimonte, M.A. 1991. On the demise of everyday memory. *American Psychologist*, 46, 27-31.

Brandimonte, M. A., Passolunghi, M. C., 1994. The effect of cue-familiarity, cue-distinctiveness, and retention interval on prospective remembering. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 47A(3), 565–587.

Brandimonte, M.A., Einstein, G.O., McDaniel, M.A., (Eds.). 1996. Prospective memory: Theory and applications. Mahwah, NJ: Erlbaum.

Brandimonte, M.A., 2005. Memory for future actions, presented at IUSOB Conference, Naples, Italy.

Brandimonte M. A., Ferrante D., Bianco C., Villani M. G., 2010. Memory for pro-social intentions: when competing motives collide. *Cognition*, 114, 436–441.

Brandimonte, M., Einstein, G., McDaniel, M. 2016. Prospective memory : Theory and applications (Paperback ed.).

Brandstätter, V., Lengfelder, A., Gollwitzer, P.M., 2001. Implementation Intentions and efficient action initiation. *Journal of Personality and Social Psychology*, 81, 946–960.

Brewer, G. A., Knight, J. B., Marsh, R. L., Unsworth, N., 2010. Individual differences in event-based PM: Evidence for multiple processes supporting cue detection. *Memory & Cognition*, 38, 304 –311.

Brom, S., Kliegel, M., Mayr, U., 2014. Improving Everyday Prospective Memory Performance in Older Adults: Comparing Cognitive Process and Strategy Training. *Psychology and Aging*, 29(3), 744-755.

Brown, W., 1923. To what extent is memory measured by a single recall?. *Journal of Experimental Psychology*, 6(5), 377.

Brown, G. D., Neath, I., Chater, N., 2007. A temporal ratio model of memory. *Psychological review*, 114(3), 539.

Brush, J.A., Camp, C.J., 1998. Using spaced retrieval as an intervention during speech- language therapy. *Clinical Gerontology*, 19 (1), 51–64.

Budson, A.E., Simons, J.S., Waring, J.D., Sullivan, A.L., Hussoin, T., Schacter, D.L., 2007. Memory for the September 11, 2001, terrorist attacks one year later in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Cortex*, 43 (7), 875–888.

Bugg, J., Scullin, M., McDaniel, M., 2013. Strengthening encoding via implementation intention formation increases prospective memory commission errors. *Psychonomic Bulletin & Review*, 20(3), 522-527.

Burgess, N., Hitch, G. J., 2006. A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55(4), 627-652.

Burke, D. M., Light, L. L., 1981. Memory and aging: The role of retrieval processes. *Psychological bulletin*, 90(3), 513.

Butler, C.R., Zeman, A., 2008. Recent insights into the impairment of memory in epilepsy: Transient epileptic amnesia, accelerated long-term forgetting and remote memory impairment. *Brain*, 131(9), 2243-2263.

Butler, C., Muhlert, N., Zeman, A., 2010. Accelerated long-term forgetting. In Forgetting (pp. 225-252). Psychology Press.

Butters, N., & Cermak, L. S. Alcoholic Korsakoff's Syndrome: An Information-Processing Approach to Amnesia. New York: Academic Press, 1980.

Butters, N., Cermak, L. S.,1986. A case study of the forgetting of autobiographical knowledge: Implications for the study of retrograde amnesia. *Autobiographical memory*, 253-272.

Callender, A. A., McDaniel, M. A., 2007. The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology*, 99(2), 339.

Camp, C.J., 2001. From efficacy to effectiveness to diffusion: making the transitions in dementia intervention research. *Neuropsychological Rehabilitation*,11, 495–517.

Camp, C.J., Foss, J.W., Stevens, A.B., O'Hanlon, A.M., 1996. Improving prospective memory task performance in persons with Alzheimer's disease. In: Brandimonte, M., Einstein, G.O., McDaniel, M.A. (Eds.), Prospective Memory: Theory and Applications. Lawrence Erlbaum, Mahwah, NJ, pp. 351–367.

Camp, C.J., Stevens, A.B., 1990. Spaced-retrieval: a memory intervention for dementia of the Alzheimer's type. *Clinical Gerontology*, 10 (1), 58–61.

Camus, J., Nicolas, S., Wenisch, E., Morrone, I., Blanchard, F., Bakchine, S., 2003. Implicit memory for words presented in short texts is preserved in Alzheimer's disease. *Psychological Medicine*, 33(1), 169-174.

Carlesimo, G.A., Sabbadini, M., Fadda, L., Caltagirone, C., 1995. Forgetting from long- term memory in dementia and pure amnesia: role of task, delay of assessment and aetiology of cerebral damage. *Cortex*, 31(2), 285–300.

Carp, J., 2012. The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, 63(1), 289-300.

Carpenter, S. K., DeLosh, E. L., 2005. Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, 19(5), 619–636.

Carpenter, S. K., DeLosh, E. L., 2006. Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34(2), 268–276.

Carpenter, S.K., Pashler, H., 2007. Testing beyond words: using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14, 474–478.

Carpenter, S. K., Pashler, H., Vul, E., 2006. What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13(5), 826–830

Carpenter, S. K., 2009. Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569.

Carpenter, S. K., 2012. Testing enhances the transfer of learning. *Current Directions in Psychological Science*, 21(5), 279–283.

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., Perfect, T., 2007. Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology,* 19(4-5), 580-606.

Cassel, A., Morris, R., Koutroumanidis, M., Kopelman, M., 2016. Forgetting in temporal lobe epilepsy: when does it become accelerated? *Cortex*, 78, 70–84.

Cavanaugh, J. C., Grady, J. G., Perlmutter, M., 1983. Forgetting and use of memory aids in 20 to 70 year olds everyday life. *The International Journal of Aging and Human Development*, 17(2), 113-122.

Cerciello, M., et al., 2017. Assessment of free and cued recall in Alzheimer's disease and vascular and frontotemporal dementia with 24-item Grober and Buschke test. *Journal of the Neurological Sciences*, 38 (1), 115–122.

Chan, J. C. K., McDermott, K. B., Roediger, H. L., III., 2006. Retrieval- induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135, 553–571.

Chan, J. C., McDermott, K. B., 2007. The testing effect in recognition memory: a dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 431- 437.

Chan, J. C., 2009. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61(2), 153-170.

Chan, J.C.K., Erdman, M.R., Davis, S.D., 2015. Retrieval induces forgetting, but only when nontested items compete for retrieval: implication for interference, inhibition, and context reinstatement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (5), 1298–1315.

Chasteen, A. L., Park, D. C., Schwarz, N., 2001. Implementation intentions and facilitation of prospective memory. *Psychological Science,* 12(6), 457-461.

Chen, X. J., Wang, Y., Liu, L. L., Cui, J. F., Gan, M. Y., Shum, D. H., Chan, R. C., 2015. The effect of implementation intention on prospective memory: A systematic and meta-analytic review. *Psychiatry Research*, 226(1), 14-22.

Cheyne, J. A., Carriere, J. S., Smilek, D., 2006. Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and cognition*, 15(3), 578-592.

Christensen, H., Kopelman, M.D., Stanhope, N., Lorentz, L., Owen, P., 1998. Rates of forgetting in Alzheimer dementia. *Neuropsychologia*, 36 (6), 547–557.

Clare, L., Wilson, B.A., Carter, G., Roth, I., Hodges, J.R., 2002. Relearning face-name associations in early Alzheimer's disease. *Neuropsychology,* 16 (4), 538–547.

Clare, L., Woods, R.T., 2004. Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: a review. *Neuropsychological Rehabilitation*, 14 (4), 385–401.

Clifasefi, S. L., Garry, M., Loftus, E. F., 2007. Setting the record (or video camera) straight on memory: The video camera model of memory and other memory myths. In S. Della Salla (Ed.), Tall tales about the mind and brain: Separating fact from fiction, 60-75.

Cohen, J. D., O'Reilly, R. C., 1996. A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. *Prospective memory: Theory and applications*, 267-295.

Cohen, A. L., West, R., Craik, F. I., 2001. Modulation of the prospective and retrospective components of memory for intentions in younger and older adults. *Aging, Neuropsychology, and Cognition*, 8(1), 1-13.

Cohen, A.L., Gollwitzer, P. M., 2008. The cost of remembering to remember: Cognitive load and implementation intentions influence ongoing task performance. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives (p. 367–390). Taylor & Francis Group/Lawrence Erlbaum Associates.

Cona, G., Arcara, G., Tarantino, V., Bisiacchi, P. S., 2012. Age-related differences in the neural correlates of remembering time-based intentions. *Neuropsychologia*, 50(11), 2692-2704.

Conroy, R., Salmon, K., 2006. Talking about parts of a past experience: The impact of discussion style and event structure on memory for discussed and nondiscussed information. *Journal of Experimental Child Psychology*, 95, 278-297.

Conte, A. M., McBride, D. M. (2018). Comparing time-based and event-based prospective memory over short delays. *Memory*, 26(7), 936-945.

Coons, E. E., Miller, N. E., 1960. Conflict versus consolidation of memory traces to explain" retrograde amnesia" produced by ECS. *Journal of comparative and physiological Psychology*, 53(6), 524.

Corkin, S., Growdon, J. H., Nissen, M. J., Huff, F. J., Freed, D. M., Sagar, H. J., 1984. Recent advances in the neuropsychological study of Alzheimer's disease. *Alzheimer's disease: Advances in basic research and therapies*, 75-93.

Craik, F. I. M., 1983. On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 302(1110), 341-359.

Craik, K. H., 1986. Personality research methods: An historical perspective. *Journal of Personality*, 54(1), 18-51.

Craik, F. I., McDowd, J. M., 1987. Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 474.

Craik, F. I., Morris, L. W., Morris, R. G., Loewen, E. R., 1990. Relations between source amnesia and frontal lobe functioning in older adults. *Psychology and aging,* 5(1), 148.

Craik, F. I., Rose, N. S., 2012. Memory encoding and aging: a neurocognitive perspective. *Neuroscience & Biobehavioral Reviews*, 36(7), 1729-1739.

Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O., De Lacey, G., 1989. Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*, 28(3), 267-273.

Crawford, J., Henry, J., Ward, A., Blake, J., 2006. The Prospective and Retrospective Memory Questionnaire (PRMQ): Latent structure, normative data and discrepancy analysis for proxy-ratings. *British Journal of Clinical Psychology,* 45(1), 83-104.

Creighton, A., Van der Ploeg, E., O'Connor, D., 2013. A literature review of spaced-retrieval interventions: a direct memory intervention for people with dementia. *International Psychogeriatrics*, 25 (11).

Crovitz, H. F., Daniel, W. F., 1984. Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, 22(5), 413-414.

Crowder, R. G., 1996. The trouble with prospective memory: A provocation. *Prospective memory: Theory and applications*, 143-147.

Cubelli, R., 2010. A new taxonomy of memory and forgetting. Forgetting (pp. 49-62). Psychology Press.

Cull, W. L., 2000. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215–235.

Cull, W.L., Shaughnessy, J.J., Zechmeister, E.B., 1996. Expanding the understanding of the expanding-pattern-of-retrieval mnemonic: toward confidence in applicability. *The Journal of Experimental Psychology*, 2, 365–378.

Davidson, M., Dorris, L., O'Regan, M., Zuberi, S.M., 2007. Memory consolidation and accelerated forgetting in children with idiopathic generalized epilepsy. *Epilepsy & Behavior*, 11(3), 394–400.

Davis, H. P., Small, S. A., Stern, Y., Mayeux, R., Feldstein, S. N., Keller, F. R., 2003. Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex*, 39(4-5), 1063-1091.

De Renzi, E., Lucchelli, F., 1993. Dense retrograde amnesia, intact learning capability and abnormal forgetting rate: A consolidation deficit?. *Cortex*, 29(3), 449-466.

Degenszajn, J., Caramelli, P., Caixeta, L., Nitrini, R., 2001. Encoding process in delayed recall impairment and rate of forgetting in Alzheimer's disease. *Arquivos de Neuro-Psiquiatria*, 59 (2A), 171–174.

Delaney, P. F., Verkoeijen, P. P. J. L., Spirgel, A., 2010. Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, 53, 63–147.

Della Sala, S. (Ed.). (2010). Forgetting. Psychology Press.

Della Sala, S., Kozlova, I., Stamate, A., Parra, M., 2018. A transcultural cognitive marker of Alzheimer's Disease. *International Journal of Geriatric Psychiatry*, 33 (6), 849–856.

Devolder, P. A., Brigham, M. C., Pressley, M., 1990. Memory performance awareness in younger and older adults. *Psychology and Aging*, 5(2), 291.

Dewar, M. T., Cowan, N., Della Sala, S., 2007. Forgetting due to retroactive interference: a fusion of early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, 43(5), 616.

Dismukes, R. K., Nowinski, J., 2007. Prospective memory, concurrent task management, and pilot error. *Attention: From theory to practice*, 4, 225.

Dismukes, R. K., 2010. Remembrance of things future: Prospective memory in laboratory, workplace, and everyday settings. *Reviews of human factors and ergonomics*, 6(1), 79-122.

Divis, K., Benjamin, A. S., 2014. Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, 42, 1049-1062.

Dobbs, A. R., Rule, B. G., 1987. Prospective memory and self-reports of memory abilities in older adults. *Canadian Journal of Psychology*, 41(2), 209–222.

Duff, K., Beglinger, L.J., Van Der Heiden, S., Moser, D.J., Arndt, S., Schultz, S.K., Paulsen, J.S., 2008. Short-term practice effects in amnestic mild cognitive impairment: implications for diagnosis and treatment. *International Psychogeriatrics*, 20 (5), 986–999.

Duff, K., Horn, K. P., Foster, N. L., Hoffman, J. M., 2015. Short-term practice effects and brain hypometabolism: preliminary data from an FDG PET study. *Archives of Clinical Neuropsychology*, 30(3), 264-270.

Duff, M. C., Covington, N. V., Hilverman, C., Cohen, N. J., 2020. Semantic memory and the hippocampus: Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in Human Neuroscience*, 13, 471.

Duncan, C. P., 1949. The retroactive effect of electroshock on learning. *Journal of comparative and physiological psychology*, 42(1), 32.

Ebbinghaus, H., 1880. Urmanuskript "Ueber das Gedächtniß". Passau: Passavia Universitätsverlag.

Ebbinghaus, H., 1885. Über das Gedächtnis. Leipzig: Dunker.

Ebbinghaus, H., 1964. Memory: A contribution to experimental psychology. Dover.

Ebbinghaus, H., 1913. Memory (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College. (Original work published in 1885; paperback ed., New York: Dover, 1964).

Einstein, G., McDaniel, M., 1990. Normal Aging and Prospective Memory. Journal of experimental psychology. *Learning, memory, and cognition*, 16. 717-26.

Einstein, G. O., Holland, L. J., McDaniel, M. A., Guynn, M. J., 1992. Age-related deficits in prospective memory: The influence of task complexity. *Psychology and aging*, 7(3), 471.

Einstein, G. O., McDaniel, M. A., Richardson, S. L., Guynn, M. J., Cunfer, A. R., 1995. Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 996.

Einstein, G. O., McDaniel, M. A., 1996. Retrieval processes in PM: Theoretical approaches and some new empirical findings. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), Prospective memory: Theory and applications (p. 115–141). Lawrence Erlbaum Associates Publishers.

Einstein, G. O., Smith, R. E., McDaniel, M. A., Shaw, P., 1997. Aging and prospective memory: The influence of increased task demands at encoding and retrieval. *Psychology and aging*, 12(3), 479.

Einstein, G. O., McDaniel, M. A., Smith, R. E., Shaw, P., 1998. Habitual PM and aging: Remembering intentions and forgetting actions. *Psychological Science*, 9, 284–288.

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., Breneiser, J., 2005. Multiple processes in prospective memory retrieval: factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General*, 134(3), 327.

Elixhauser, A., Leidy, N. K., Meador, K., Means, E., Willian, M. K., 1999. The relationship between memory performance, perceived cognitive function, and mood in patients with epilepsy. *Epilepsy research*, 37(1), 13-24.

Elliot, G., Isaac, C.L., Muhlert, N., 2014. Measuring forgetting: a critical review of accelerated long-term forgetting studies. *Cortex*, 54, 16–32.

Ellis, J. A., Nimmo-Smith, I., 1993. Recollecting naturally-occurring intentions: A study of cognitive and affective factors. *Memory*, 1(2), 107-126.

Ellis, J., 1996. Prospective memory or the realization of delayed intentions: A conceptual framework for research. *Prospective memory: Theory and applications*, 1-22.

Ellis, J., Kvavilashvili, L., 2000. Prospective memory in 2000: Past, present, and future directions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(7), S1-S9.

Esopenko, C., Levine, B., 2017. Autobiographical memory and structural brain changes in chronic phase TBI. *Cortex*, 89, 1-10.

Farrell, S., Lewandowsky, S., 2002. An endogenous distributed model of ordering in serial recall. *Psychonomic bulletin & review*, 9(1), 59-79.

Ferrer, E., Salthouse, T., Stewart, W., Schwartz, B., 2004. Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging*, 19 (2), 243–259.

Finn, B., 2017. A Framework of Episodic Updating: An Account of Memory Updating After Retrieval. *Psychology of Learning and Motivation*, 67, 173–211.

Fitzgerald, Z., Thayer, Z., Mohamed, A., Miller, L. A., 2013. Examining factors related to accelerated long-term forgetting in epilepsy using ambulatory EEG monitoring. *Epilepsia*, 54(5), 819-827.

Fjell, A. M., Walhovd, K. B., Reinvang, I., Lundervold, A., Dale, A. M., Quinn, B. T., ... Fischl, B., 2005. Age does not increase rate of forgetting over weeks—neuroanatomical volumes and visual memory across the adult life-span. *Journal of the International Neuropsychological Society*, 11(1), 2-15.

Folstein, Marshal F., Folstein, Susan E., Mchugh, Paul R. 1975. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198.

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., Benjamin, A. S., 2019. Aging and recognition memory: A meta-analysis. *Psychological bulletin*, 145(4), 339.

Freed, D. M., Corkin, S., Cohen, N. J., 1987. Forgetting in HM: A second look. *Neuropsychologia*, 25(3), 461-471.

Freed, D.M., & Corkin, S., 1988. Rate of forgetting in H.M.: 6-month recognition. *Behavioral Neuroscience*, 102 (6), 823–827.

Freed, D.M., Corkin, S., & Cohen, N.J. 1987. Forgetting in H.M.: a second look. *Neuropsychologia*, 25(3), 461-71.

Friedman, W., & Steinberg, Robert J., 1993. Memory for the Time of Past Events. *Psychological Bulletin*, 113(1), 44-66.

Frisk, V., & Milner, B., 1990a. The relationship of working memory to the immediate recall of stories following unilateral temporal or frontal lobectomy. *Neuropsychologia*, *28*(2), 121-135.

Frisk, V., Milner, B., 1990b. The role of the left hippocampal region in the acquisition and retention of story content. *Neuropsychologia*, 28(4), 349-359.

Frisk, V., & Milner, B., 1991. Does left temporal lobectomy adversely affect the rate at which verbal material can be processed?. *Neuropsychologia*, *29*(2), 113-123.

Garcia-Bajos, E., Migueles, M., Anderson, M. C., 2009. Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory*, 17(1), 92-103.

García-Bajos, E., Migueles, M., 2013. Script-driven processing affords protection from retrieval-induced forgetting in the recall of everyday activities. *The Quarterly Journal of Experimental Psychology*, 66(7), 1317-1330.

Garin, O., 2014. Ceiling effect. In: Michalos, A.C. (Ed.), Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.

Gates, A. I., 1917. Recitation as a factor in memorizing. *Archives of Psychology*, 6(40). 181–206.

Gauld, A., Stephenson, G. M., 1967. Some experiments relating to Bartlett's theory of remembering. *British Journal of Psychology*, 58(1-2), 39-49.

Geffen, G.M., Geffen, L., Bishop, K., Manning, L., 1997. Extended delayed recall of AVLT word lists: effects of age and sex on adult performance. *Australian Journal of Psychology*, 49(2), 78-84.

Gelman, A., Loken, E., 2014. The statistical crisis in science: data-dependent analysis--a" garden of forking paths"-explains why many statistically significant comparisons don't hold up. *American scientist*, 102(6), 460-466.

Gelman, A., Loken, E. 2016. The statistical crisis in science. *The best writing on mathematics*, *2015*, 305.

Gentile, J. R., Monaco, N., Iheozor-Ejiofor, I. E., Ndu, A. N., Ogbonaya, P. K., 1982. Retention by "fast" and "slow" learners. *Intelligence*, 6, 125-138.

Gentile, J.R., Voelkl, K.E, Pleasant, J.M., Monaco, N.M., 1995. *The Journal of Experimental Education*, 63 (3), 185-197.

Geurts, S., van Der Werf, S.P., Kessels, R.P.C., 2015. Accelerated forgetting? An evaluation on the use of long-term forgetting rates in patients with memory problems. *Frontiers in Psychology*, 6(752), 752.

Gilbert, S. J., 2011. Decoding the content of delayed intentions. *Journal of Neuroscience*, 31(8), 2888-2894.

Gillette, A. L., 1936. Learning and retention: A comparison of three experimental procedures. *Archives of psychology*, 28(198), 5–56.

Gold, P. E., 2006. The many faces of amnesia. *Learning & Memory*, 13(5), 506-514.

Goldberg, T.E., Harvey, P.D., Wesnes, K.A., Snyder, P.J., Schneider, L.S., 2015. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer’s disease randomized controlled trials. *Alzheimer’s Dementia*, 1(1), 103–111.

Gollwitzer, P. M., 1993. Goal achievement: The role of intentions. *European review of social psychology*, 4(1), 141-185.

Gollwitzer, P. M., & Brandstätter, V., 1997. Implementation intentions and effective goal pursuit. *Journal of personality and social psychology*, 73(1), 186.

Gollwitzer, P. M., 1999. Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54(7), 493–503.

Gollwitzer, P., Sheeran, P., 2006. Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes. *Advances in Experimental Social Psychology*, (38), 69-119.

Goschke, T., Kuhl, J., 1993. Representation of intentions: Persisting activation in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1211–1226.

Goschke, T., Kuhl, J., 1996. Remembering what to do: Explicit and implicit memory for intentions. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (p. 53–91). Lawrence Erlbaum Associates Publishers.

Graf, P., 2001. Prospective Memory: A New Focus for Research. *Consciousness and Cognition*, 10, 437–450.

Greene, J.D.W., Baddeley, A.D., Hodges, J.R., 1996. Analysis of the episodic memory deficit in early Alzheimer’s Disease: evidence from the doors and people test. *Neuropsychologia*, 34, 537–551.

Green, R.E., Kopelman, M.D., 2002. Contribution of recollection and familiarity judgements to rate of forgetting in organic amnesia. *Cortex*, 38 (2), 161–178.

Greving, S., Richter, T., 2018. Examining the testing effect in university teaching: retrievability and question format matter. *Frontiers in psychology*, 9, 2412.

Grober, E., Kawas, C., 1997. Learning and retention in preclinical and early Alzheimer's disease. *Psychology and Aging*, 12(1), 183–188.

Guynn, M. J., McDaniel, M. A., Einstein, G. O., 1998. Prospective memory: When reminders fail. *Memory and cognition*, 26(2), 287-298.

Halamish, V., Bjork, R. A., 2011. When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37, 801–812.

Hardt, O., Einarsson, E. Ö., Nader, K., 2010. A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual review of psychology*, 61, 141-167.

Hardt, O., Nader, K., Nadel, L., 2013. Decay happens: the role of active forgetting in memory. *Trends in cognitive sciences*, 17(3), 111-120.

Harker, J. O., Riege, W. H., 1985. Aging and delay effects on recognition of words and designs. *Journal of Gerontology*, 40(5), 601-604.

Harris, J. E., 1984. Remembering to do things: A forgotten topic. In J. E. Harris & P. E. Morris (Eds.), Everyday memory actions and absent-mindedness (pp. 71-92). New York: Academic Press.

Hart, R.P., Kwentus, J.A., Harkins, S.W., Taylor, J.R., 1988. Rate of forgetting in mild Alzheimer's-type dementia. *Brain and Cognition*, 7(1), 31–38.

Hart, R.P., Kwentus, J.A., Taylor, J.R., Harkins, S.W., 1987. Rate of forgetting in dementia and depression. *Journal of Consulting and Clinical Psychology*, 55 (1), 101.

Harwood, E., Naylor, G. F. K., 1969. Recall and recognition in elderly and young subjects. *Australian Journal of Psychology*, 21(3), 251-257.

Hashtroudi, S., Johnson, M., Chrosniak, L., 1989. Aging and source monitoring. *Psychology and Aging*, 4(1), 106-12.

Hay, J., Jacoby, L., Light, Leah L., 1999. Separating Habit and Recollection in Young and Older Adults: Effects of Elaborative Processing and Distinctiveness. *Psychology and Aging*, 14(1), 122-134.

Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., Grant, I., 2001. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of clinical neuropsychology*, 16(1), 75-91.

Henry, J. D., MacLeod, M. S., Phillips, L. H., Crawford, J. R., 2004. A meta-analytic review of prospective memory and aging. *Psychology and aging*, 19(1), 27.

Herrmann, D., Miller, George A., 1982. Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, 92(2), 434-452.

Hertzog, C., Dixon, R. A., Hultsch, D. F., 1990. Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and aging*, 5(2), 215.

Hertzog, C., Dixon, R. A., 1994. Metacognitive development in adulthood and old age. *Metacognition: Knowing about knowing*, 227-251.

Hicks, J. L., Marsh, R. L., Russell, E. J., 2000. The properties of retention intervals and their affect on retaining prospective memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1160.

Holbrook, J., Dismukes, K., 2009. Prospective memory in everyday tasks. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 10, pp. 590-594). Sage CA: Los Angeles, CA: SAGE Publications.

Hornberger, M., Piguet, O., 2012. Episodic memory in frontotemporal dementia: a critical review. *Brain*, 135(3), 678-692.

Hulbert, J. C., Shivde, G., Anderson, M. C., 2012. Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. *Experimental Psychology*, 59, 11–21.

Hulicka, I. M., Rust, L. D., 1964. Age-related retention deficit as a function of learning. *Journal of the American Geriatrics Society*, 12(11), 1061-1065.

Hupbach, A., Gomez, R., Hardt, O., Nadel, L., 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learning & memory*, 14(1-2), 47-53.

Huppert, F. A., Piercy, M., 1977. Recognition memory in amnesic patients: a defect of acquisition?. *Neuropsychologia*, 15(4-5), 643-652.

Huppert, F. A., Piercy, M., 1978. Dissociation between learning and remembering in organic amnesia. *Nature*, 275(5678), 317-318.

Huppert, F. A., Piercy, M., 1979. Normal and abnormal forgetting in organic amnesia: effect of locus of lesion. *Cortex*, 15(3), 385-390.

Huppert, F. A., Kopelman, M. D., 1989. Rates of forgetting in normal ageing: A comparison with dementia. *Neuropsychologia*, 27(6), 849-860.

Inglis, J., 1959. 'A paired-associate learning test for use with elderly psychiatric patients.' The Journal of mental science, 105, 440–3.

Isaac, C.L., Mayes, A.R., 1999a. Rate of forgetting in amnesia: I. Recall and recognition of prose. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 942–962.

Isaac, C.L., Mayes, A.R., 1999b. Rate of forgetting in amnesia: II. Recall and recognition of word lists at different levels of organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 963–977.

Jansari, A.S., Davis, K., McGibbon, T., Firminger, S., Kapur, N., 2010. When "long-term memory" no longer means "forever": analysis of accelerated long-term forgetting in a patient with temporal lobe epilepsy. *Neuropsychologia*, 48(6), 1707–1715.

Java, R. I., Gardiner, J. M., 1991. Priming and aging: Further evidence of preserved memory function. *The American Journal of Psychology*, 89-100.

Jenkins, J. G., Dallenbach, K. M., 1924. Obliviscence during sleep and waking. *The American Journal of Psychology*, 35(4), 605-612.

Jennings, J., Jacoby, L., 1993. Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, 8(2), 283-93.

Johnson, M., Raye, C., Estes, William K., 1981. Reality monitoring. *Psychological Review*, 88(1), 67-85.

Johnson, R. E., 1962. The retention of qualitative changes in learning. *Journal of Verbal Learning & Verbal Behavior*, 1, 218-223.

Jonker, T. R., Seli, P., MacLeod, C. M., 2013. Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review*, 120, 852–872.

Kandel, E. R., 2001. The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, 294(5544), 1030-1038.

Kang, S. H. K., McDermott, K. B., Roediger, H. L. III., 2007. Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528–558.

Kapur, N., Scholey, K., Moore, E., Barker, S., Brice, J., Thompson, S., ... Fleming, J., 1996. Long-term retention deficits in two cases of disproportionate retrograde amnesia. *Journal of Cognitive Neuroscience*, 8(5), 416-434.

Kapur, N., Millar, J., Colbourn, C., Abbott, P., Kennedy, P., Docherty, T., 1997. Very long-term amnesia in association with temporal lobe epilepsy: evidence for multiple-stage consolidation processes. *Brain and Cognition*, 35(1), 58-70.

Karpicke, J. D., Roediger III, H. L., 2007. Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151-162.

Karpicke, J. D., Roediger III, H. L., 2008. The critical importance of retrieval for learning. *Science*, 319, 966-968.

Karpicke, J. D., 2009. Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469–486.

Karpicke, J. D., Lehman, M., Aue, W. R., 2014. Retrieval-based learning: An episodic context account. *The Psychology of Learning and Motivation*, 61, 237–284.

Karpicke, J. D., 2017. Retrieval-Based Learning: A Decade of Progress. Grantee Submission.

Katz, S., 1983. Assessing self-maintenance: activities of daily living, mobility and instrumental activities of daily living. *Journal of the American Geriatrics Society*, 31(12), 721–726.

Kausler, D. H., & Wiley, J. G. (1991). Effects of short-term retrieval on adult age differences in long-term recall of actions. *Psychology and aging*, 6(4), 661.

Kausler, D., Hakami, M., Scarr, S., 1983. Memory for activities: Adult age differences and intentionality. *Developmental Psychology*, 19(6), 889-894.

Kausler, D. H., 1994. Learning and memory in normal aging (pp. 276-305). San Diego, CA: Academic Press.

Kausler, D., Lichty, W., Davis, R., Scarr, Sandra., 1985. Temporal Memory for Performed Activities: Intentionality and Adult Age Differences. *Developmental Psychology*, 21(6), 1132-1138.

Keppel, G., Underwood, B. J., 1962. Proactive inhibition in short-term retention of single items. *Journal of verbal learning and verbal behavior,* 1(3), 153-161.

Wickens, T. D., Keppel, G., 2004. Design and analysis: A researcher's handbook. Pearson Prentice-Hall.

Kim, H., Cabeza, R., 2007. Trusting our memories: Dissociating the neural correlates of confidence in veridical versus illusory memories. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(45), 12190-7.

Kinsella, G.J., Ong, B., Storey, E., Wallace, J., Hester, R., 2007. Elaborated spaced-retrieval and prospective memory in mild Alzheimer's disease. *Neuropsychological Rehabilitation*, 17(6), 688–706.

Kliegel, M., Martin, M., 2003. Prospective memory research: Why is it relevant?. *International journal of psychology,* 38(4), 193-194.

Kliegel, M., Jager, T., 2006. Delayed–execute prospective memory performance: The effects of age and working memory. *Developmental neuropsychology,* 30(3), 819-843.

Kliegel M., Martin M., McDaniel M. A., Einstein G. O., 2001. Varying the importance of a PM task: differential effects across time and event-based PM. *Memory,* 9, 1–11.

Kliegel, M., Martin, M., 2003. Prospective memory research: Why is it relevant?. *International journal of psychology*, 38(4), 193-194.

Kliegel M., Martin M., McDaniel M. A., Einstein G. O., 2004. Importance effects on performance in event-based PM tasks. *Memory*, 12, 553–561.

Kliegel, M., Jager, T., 2006. Delayed–execute prospective memory performance: The effects of age and working memory. *Developmental neuropsychology*, 30(3), 819-843.

Kliegel, M., McDaniel, M. A., Einstein, G. O. (Eds.). 2007. Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives. Psychology Press.

Kliegel, M., McDaniel, M., Einstein, G., 2008. Prospective memory : Cognitive, neuroscience, developmental, and applied perspectives. New York ; London: Lawrence Erlbaum Associates.

Kliegel, M., Ballhausen, N., Hering, A., Ihle, A., Schnitzspahn, K., Zuber, S., 2016. Prospective Memory in Older Adults: Where We Are Now and What Is Next. *Gerontology*, 62(4), 459-466.

Knight, J. B., Meeks, J. T., Marsh, R. L., Cook, G. I., Brewer, G. A., Hicks, J. L., 2011. An observation on the spontaneous noticing of prospective memory event-based cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 298.

Koechlin, E., Hyafil, A., 2007. Anterior prefrontal function and the limits of human decision-making. *Science*, 318(5850), 594-598.

Kopelman, M.D., 1985. Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia*, 23, 623–638.

Kopelman, M. D., 1992. 3 Storage, Forgetting, and Retrieval in the Anterograde and Retrograde Amnesia of Alzheimer Dementia. In Advances in psychology (Vol. 89, pp. 45-71). North-Holland.

Kopelman, M.D., Stanhope, N., 1997. Rates of forgetting in organic amnesia following temporal lobe, diencephalic, or frontal lobe lesions. *Neuropsychology*, 11 (3), 343–356.

Koriat, Ben-Zur, Sheffer., 1988. Telling the same story twice: Output monitoring and age. *Journal of Memory and Language*, 27(1), 23-39.

Kornell, N., Bjork, R. A., Garcia, M. A., 2011. Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97.

Kornell, N., Vaughn, K. E., 2016. How retrieval attempts affect learning: A review and synthesis. In B. H. Ross (Ed.), Psychology of learning and motivation (Vol. 65, pp. 183–212). Amsterdam, The Netherlands: Elsevier.

Kramer, J. H., Schuff, N., Reed, B. R., Mungas, D., Du, A. T., Rosen, H. J., ... Chui, H. C, 2004. Hippocampal volume and retention in Alzheimer's disease. *Journal of the International Neuropsychological Society*, 10(4), 639-643.

Kvavilashvili, L., 1987. Remembering intention as a distinct form of memory. *British journal of psychology*, 78(4), 507-518.

Kvavilashvili, L., 1992. Remembering intentions: A critical review of existing experimental paradigms. *Applied Cognitive Psychology*, 6(6), 507-524.

Kvavilashvili, L., Ellis, J., 1996. Varieties of intention: Some distinctions and classifications. *Prospective memory: Theory and applications*, 6, 183-207.

Kvavilashvili, L., Fisher, L., 2007. Is time-based prospective remembering mediated by self-initiated rehearsals? Role of incidental cues, ongoing activity, age, and motivation. *Journal of Experimental Psychology: General*, 136(1), 112.

Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., Milne, A., 2009. Differential effects of age on prospective and retrospective memory tasks in young, young-old, and old-old adults. *Memory*, 17(2), 180-196.

Kyllonen, P. C., Tirre, W. C., 1988. Individual differences in associative learning and forgetting. *Intelligence*, 12, 393–421.

Lah, S., Black, C., Gascoigne, M. B., Gott, C., Epps, A., Parry, L., 2017. Accelerated long-term forgetting is not epilepsy specific: evidence from childhood traumatic brain injury. *Journal of neurotrauma*, 34(17), 2536-2544.

Larrabee, G., Youngjohn, J., Sudilovsky, A., Crook, T., 1993. Accelerated forgetting in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology*, 15(5), 701–712.

Lawton, M.P., Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontology*, 9, 179–186.

Leeming, F. C., 2002. The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, 29(3), 210-212.

Leitz, J. R., Morgan, C. J., Bisby, J. A., Rendell, P. G., & Curran, H. V. (2009). Global impairment of prospective memory following acute alcohol. *Psychopharmacology*, 205(3), 379-387.

Lemaire, B., Portrat, S., 2018. A computational model of working memory integrating time-based decay and interference. *Frontiers in psychology*, 9, 416.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). Neuropsychological assessment. Oxford University Press, USA.

Lim, C., Alexander, M. P., 2009. Stroke and episodic memory disorders. *Neuropsychologia*, 47(14), 3045-3058.

Little, J. L., Storm, B. C., Bjork, E. L., 2011. The costs and benefits of testing text materials. *Memory*, 19(4), 346-359.

Loft, S., Yeo, G., 2007. An investigation into the resource requirements of event-based prospective memory. *Memory & Cognition*, 35(2), 263-274.

Loftus, G. R., 1985a. Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 397-406.

Loftus, G. R., 1985b. Consistency and confoundings: Reply to Slamecka. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 817-820.

Loftus, E. F., 1971. Memory for intentions: The effect of presence of a cue and interpolated activity. *Psychonomic Science*, 23(4), 315-316.

Logie, R. H., Maylor, E. A., 2009. An Internet study of prospective memory across adulthood. *Psychology and aging*, 24(3), 767.

Logie, R. H., Wolters, M., Niven, E., 2018. Preserving and forgetting in the human brain. In Personal Multimedia Preservation (pp. 9-45). Springer, Cham.

Lombardi, M., Perri, R., Fadda, L., Caltagirone, C., Carlesimo, G., 2018. Forgetting of the recollection and familiarity components of recognition in patients with amnestic mild cognitive impairment. *Journal of Neuropsychology*, 12(2), 231–247.

Lucchelli, F., Spinnler, H., 1998. Ephemeral new traces and evaporated remote engrams: a form of neocortical temporal lobe amnesia? A preliminary case report. *Neurocase*, 4(6), 447–459.

Luh, C. W., 1922. The conditions of retention. *Psychological Monographs*, 31(3), i.

Lustig, C., Buckner, R. L., 2004. Preserved neural correlates of priming in old age and dementia. *Neuron*, 42(5), 865-875.

MacDonald, S.W.S, Derwinger, A., Stigsdotter-Neely, A., Backman, L., 2006. Rate of Acquisition, Adult Age, and Basic Cognitive Abilities Predict Forgetting: New Views on a Classic Problem. *Journal of Experimental Psychology*, 135(3), 368–390.

MacLeod, M. D., Macrae, C. N., 2001. Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, 12(2), 148-152.

MacLeod, M. D., Hulbert, J. C., 2011. Sleep, retrieval inhibition, and the resolving power of human memory. In Successful Remembering and Successful Forgetting (pp. 151-170). Psychology Press.

Mameniskiene, R., Jatuzis, D., Kaubrys, G., Budrys, V. (2006). The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. *Epilepsy & Behavior,* 8(1), 278-288.

Manes, F., Graham, K.S., Zeman, A., de Lujan Calcagno, M., Hodges, J.R., 2005. Autobiographical amnesia and accelerated forgetting in transient epileptic amnesia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76(10), 1387–1391.

Manes, F., Serrano, C., Calcagno, M.L., Cardozo, J., Hodges, J., 2008. Accelerated forgetting in subjects with memory complaints. A new form of mild cognitive impairment? *Journal of Neurology*, 255, 1067–1070.

Mantyla, T., 2003. Assessing absentmindedness: Prospective memory complaint and impairment in middle-aged adults.(Brief Article). *Memory & Cognition*, 31(1), 15-25.

Marsh, R. L., Hicks, J. L., Bink, M. L., 1998. Activation of completed, uncompleted, and partially completed intentions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 350–361.

Marsh R. L., Hicks J. L., Cook G. I., Hansen J. S., Pallos A. L., 2003. Interference to ongoing activities covaries with the characteristics of an event-based intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 861–870.

Marsh, R. L., Hicks, J. L., Cook, G. I., 2006. Task interference from prospective memories covaries with contextual associations of fulfilling them. *Memory & Cognition*, 34(5), 1037-1045.

Marsh, R. L., Hicks, J. L., Cook, G. I., Mayhorn, C. B., 2007. Comparing older and younger adults in an event-based prospective memory paradigm containing an output monitoring component. *Aging, Neuropsychology, and Cognition*, 14(2), 168-188.

Martin, B. A., Brown, N. L., & Hicks, J. L. (2011). Ongoing task delays affect prospective memory more powerfully than filler task delays. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(1), 48.

Martin, B. A., Brown, N. L., Hicks, J. L., 2011. Ongoing task delays affect prospective memory more powerfully than filler task delays. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(1), 48.

Martone, M., Butters, N., Trauner, D., 1986. Some analyses of forgetting of pictorial material in amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology*, 8(3), 161-178.

Mary, A., Schreiner, S., Peigneux, P., 2013. Accelerated long-term forgetting in aging and intra-sleep awakenings. *Frontiers in Psychology*, 4, 750.

Mayer, R. E., 1983. Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology*, 75, 40-49.

Mayes, A.R., 1986. Learning and memory disorders and their assessment. *Neuropsychologia*, 24, 25–39.

Mayes, A. R., Isaac, C. L., Holdstock, J. S., Cariga, P., Gummer, A., Roberts, N., 2003. Long-term amnesia: a review and detailed illustrative case study. *Cortex*, 39(4-5), 567-603.

Maylor, E. A., 1990. Age and prospective memory. *The Quarterly Journal of Experimental Psychology*, 42(3), 471-493.

Maylor, E. A., & Logie, R. H., 2010. A large-scale comparison of prospective and retrospective memory development from childhood to middle age. *The Quarterly Journal of Experimental Psychology*, 63(3), 442-451.

McCabe, D. P., Geraci, L. D., 2009. The influence of instructions and terminology on the accuracy of remember–know judgments. *Consciousness and Cognition*, 18(2), 401-413.

McDaniel, M. A., Einstein, G. O., 1992. Aging and prospective memory: Basic findings and practical applications. *Advances in learning and behavioral disabilities*, 7(87), 105.

McDaniel, M. A., Einstein, G. O., 1993. The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory*, 1(1), 23-41.

McDaniel, M., Einstein, G., 2000. Strategic and Automatic Processes in Prospective Memory Retrieval: A Multiprocess Framework. *Applied Cognitive Psychology*. 14, S127 - S144.

McDaniel, M. A., Guynn, M. J., Einstein, G. O., Breneiser, J., 2004. Cue-focused and reflexive-associative processes in prospective memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 605.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., Morrisette, N., 2007. Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.

McDaniel, M.A., Howard, D.C. Butler, K.M., 2008. Implementation intentions facilitate PM under high attention demands. *Memory & Cognition*, 36, 716–724.

McDaniel, M. A., Bugg, J. M., Ramuschkat, G. M., Kliegel, M., Einstein, G. O., 2009. Repetition errors in habitual prospective memory: Elimination of age differences via complex actions or appropriate resource allocation. *Aging, Neuropsychology, and Cognition*, 16(5), 563-588.

McDaniel, M., Scullin, A., 2010. Implementation intention encoding does not automatize prospective memory responding. *Memory and Cognition*, 38(2), 221-232.

McFarland, C., Glisky, E. L., 2012. Implementation intentions and imagery: Individual and combined effects on PM among young adults. *Memory and Cognition*, 40(1), 62-69.

McGaugh, J. L., 2000. Memory-a century of consolidation. *Science*, 287(5451), 248-251.

McGeoch, J. A., 1932. Forgetting and the law of disuse. *Psychological review*, 39(4), 352.

McGeoch, J. A., Irion, A. L. 1952. The psychology of human learning. New York: Longmans, Green.

McIntyre, J. S., Craik, F. I., 1987. Age differences in memory for item and source information. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 41(2), 175.

McKee, R. D., Squire, L. R., 1992. Equivalent forgetting rates in long-term memory for diencephalic and medial temporal lobe amnesia. *Journal of Neuroscience*, 12(10), 3765-3772.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer’s disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer’s disease. *Neurology*, 34(7), 939–944.

Meacham, J. A., Dumitru, J., 1975. Prospective remembering and external retrieval cues.

Meacham, J. A., Leiman, B., 1982. Remembering to perform future actions." In U. Neisser (Ed.), Memory observed: Remembering in natural contexts. San Francisco: Freeman.

Meacham, J., Singer, J., 1977. Incentive Effects in Prospective Remembering. *Journal of Psychology*, 97, 191-197.

Meeks, J. T., Marsh, R., 2009. Implementation intentions about nonfocal event-based PM tasks. *Psychological research*, 74, 82-9.

Meeter, M., Murre, J. M. J., Janssen, S. M. J., 2005. Remembering the news: Modelling retention data from a study with 14,000 participants. *Memory & Cognition*, 33, 793– 810.

Meier, B., Zimmermann, T. D., Perrig, W. J., 2006. Retrieval experience in prospective memory: Strategic monitoring and spontaneous retrieval. *Memory*, 14(7), 872-889.

Meier, B., Matter, S., Baumann, B., Walter, S., & Koenig, T., 2014. From episodic to habitual prospective memory: ERP-evidence for a linear transition. *Frontiers in Human Neuroscience,* 8(918), 136-149.

Meier, B., & Zimmermann, T. D., 2015. Loads and loads and loads: the influence of prospective load, retrospective load, and ongoing task load in prospective memory. *Frontiers in human neuroscience*, 9, 322.

Meier, B., 2019. Toward an ecological approach to prospective memory? The impact of Neisser's seminal talk on prospective memory research. *Frontiers in psychology*, 10, 1005.

Midorikawa, A. Kawamura, M., 2007. Recovery of long-term anterograde amnesia, but not retrograde amnesia, after initiation of an anti-epileptic drug in a case of transient epileptic amnesia. *Neurocase*, 13, 385-389.

Migueles, M., García-Bajos, E., 2007. Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(9), 1157-1172.

Miller, G. A., Galanter, E., Pribram, K. H., 1960. Plans and the structure of behavior. New York, NY, US: Henry Holt and Co.

Minett, T. S. C., Da Silva, R. V., Ortiz, K. Z., Bertolucci, P. H. F., 2008. Subjective memory complaints in an elderly sample: a cross-sectional study. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, 23(1), 49-54

Mitchell, D. B., Brown, A. S., Murphy, D. R., 1990. Dissociations between procedural and episodic memory: effects of time and aging. *Psychology and aging*, 5(2), 264.

Mitterschiffthaler, M., Fu, C., Dalton, J., Andrew, C., Williams, S., 2007. A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, 28(11), 1150-1162.

Modigliani, V., 1976. Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 609-622.

Moscovitch, M., 1982. A neuropsychological approach to perception and memory in normal and pathological aging. In Aging and cognitive processes (pp. 55-78). Springer, Boston, MA.

Moss, M.B., Albert, M.S., Butters, N., Payne, M., 1986. Differential patterns of memory loss among patients with Alzheimer's disease, Huntington's disease and alcoholic Korsakoff s syndrome. *Archives of neurology*, 43, 239–246.

Mozer, M. C., Howe, M., Pashler, H., 2004. Using testing to enhance learning: A comparison of two hypotheses. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the twenty-sixth annual conference of the cognitive science society (pp. 975–980). Mahwah, NJ: Erlbaum.

Mueller, J. H., Brown, S. C., 1977. Output interference and intralist repetition in free recall. *American Journal of Psychology*, 90, 157-164.

Müller, G. E., Pilzecker, A., 1900. Experimentelle Beiträge zur Lehre vom Gedächtniss. *Zeitschrift für Psychologie*, 1, 1-300.

Murayama, K., Miyatsu, T., Buchli, D., Storm, B.C., 2014. Forgetting as a consequence of retrieval: a meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, 140(5), 1383-1409.

Murre, J. M., Dros, J., 2015. Replication and analysis of Ebbinghaus' forgetting curve. *PloS one*, 10(7).

Nadel, L., Moscovitch, M., 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, 7(2), 217-227.

Nader, K., Hardt, O., 2009. A single standard for memory: the case for reconsolidation. *Nature Reviews Neuroscience*, 10(3), 224-234.

Nairne, J. S., 1990. A feature model of immediate memory. *Memory & Cognition*, 18(3), 251-269.

Narayanan, J., Duncan, R., Greene, J., Leach, J. P., Razvi, S., McLean, J., Evans, J. J. 2012. Accelerated long-term forgetting in temporal lobe epilepsy: Verbal, nonverbal and autobiographical memory. *Epilepsy & Behavior*, 25(4), 622-630.

Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53 (4), 695–699.

Neath, I., Brown, G. D., 2012. Arguments against memory trace decay: A SIMPLE account of Baddeley and Scott. *Frontiers in Psychology*, 3, 35.

Nelson, H. E., 1982. National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual. Nfer-Nelson.

Nelson, S. M., Savalia, N. K., Fishell, A. K., Gilmore, A. W., Zou, F., Balota, D. A., McDermott, K. B., 2016. Default mode network activity predicts early memory decline in healthy young adults aged 18–31. *Cerebral Cortex*, 26, 3379–3389.

Neroni, M. A., Gamboz, N., Brandimonte, M. A., 2014. Does episodic future thinking improve prospective remembering?. *Consciousness and cognition*, 23, 53-62.

Neumann, O., Klotz, W., 1994. Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification. *Attention and performance XV: Conscious and nonconscious information processing*, 123-150.

Nickerson, R. S., 1984. Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & cognition*, 12(6), 531-552.

Nigro, G., Cicogna, P. C., 2000. Does delay affect prospective memory performance?. *European psychologist*, 5(3), 228.

Nilsson, L.G., Cohen, R. L., Nyberg, L., 1989. Recall of enacted and nonenacted instructions compared: Forgetting functions. *Psychological Research*, 51, 188–193.

Papies, E. K., Aarts, H., de Vries, N. K., 2009. Planning is for doing: Implementation intentions go beyond the mere creation of goal-directed associations. *Journal of Experimental Social Psychology*, 45(5), 1148-1151.

Paraskevaides, T., Morgan, C. J., Leitz, J. R., Bisby, J. A., Rendell, P. G., Curran, H. V., 2010. Drinking and future thinking: acute effects of alcohol on prospective memory and future simulation. *Psychopharmacology*, 208(2), 301.

Park, D. C., Puglisi, J. T., Smith, A. D., 1986. Memory for pictures: Does an age-related decline exist?. *Psychology and Aging*, 1(1), 11.

Park, D. C., Royal, D., Dudley, W., Morrell, R., 1988. Forgetting of pictures over a long retention interval in young and older adults. *Psychology and Aging*, 3(1), 94.

Pastötter, B., Schicker, S., Niedernhuber, J., Bauml, K. H., 2011. Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 287–297.

Pastötter, B., Bäuml, K. H. T., 2014. Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in psychology*, 5, 286.

Patton, G. W., Meit, M., 1993. Effect of aging on prospective and incidental memory. *Experimental Aging Research*, 19(2), 165-176.

Perner, J., Ruffman, T., 1995. Episodic memory and autonoetic conciousness: developmental evidence and a theory of childhood amnesia. *Journal of experimental child psychology*, 59(3), 516-548.

Petersen, R. C., Smith, G. E., Ivnik, R. J., Kokmen, E., Tangalos, E. G., 1994. Memory function in very early Alzheimer's disease. *Neurology*, 44(5), 867-867.

Piazzini, A., Canevini, M. P., Maggiori, G., Canger, R., 2001. The perception of memory failures in patients with epilepsy. *European Journal of Neurology*, 8(6), 613-620.

Pike, K. E., Savage, G., 2008. Memory profiling in mild cognitive impairment: can we determine risk for Alzheimer's disease?. *Journal of neuropsychology*, 2(2), 361-372.

Pilotti, M., Chodorow, M., Petrov, R., 2009. The usefulness of retrieval practice and review- only practice for answering conceptually related test questions. *The Journal of General Psychology*,136, 179–203.

Pink, J., Dodson, E., 2013. Negative prospective memory: Remembering not to perform an action. *Psychonomic Bulletin & Review*, 20(1), 184-190.

Piolino, P., Chételat, G., Matuszewski, V., Landeau, B., Mézenge, F., Viader, F., ... Desgranges, B., 2007. In search of autobiographical memories: A PET study in the frontal variant of frontotemporal dementia. *Neuropsychologia*, 45(12), 2730-2743.

Postman, L., Rau, L., 1957. Retention as a function of the method of measurement. University of California Publications in Psychology.

Pyc, M.A., Rawson, K.A., 2007. Examining the efficiency of schedules of distributed retrieval practice. *Memory and Cognition*. 35, 1917–1927.

Pyc, M. A., Rawson, K. A., 2010. Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335.

Quillian, M. R., 1968. Semantic memory. M Minsky, Ed, Semantic Information Processing.

Raaijmakers, J. G. W., Jakab, E., 2013. Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, 68, 98–122.

Rabbitt, P., 1996. Why are studies of “prospective memory” planless. *Prospective memory: Theory and applications*, 239-248.

Radvansky, G. A., 2015. Human memory. Psychology Press.

Raven, J., 1962. Advanced Progressive Matrices Set II.London: H.K. Lewis.

Rawson, K. A., Dunlosky, J., Thiede, K. W., 2000. The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28, 1004-1010.

Reed, B. R., Paller, K., Mungas, D., 1998. Impaired acquisition and rapid forgetting of patterned visual stimuli in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 20(5), 738-749.

Reitman, J. S., 1971. Mechanisms of forgetting in short-term memory. *Cognitive Psychology*, 2(2), 185-195.

Rendell, P. G., Thomson, D. M., 1993. The effect of ageing on remembering to remember: An investigation of simulated medication regimens. *Australian Journal on Ageing*, 12(1), 11-18.

Rendell, P. G., Thomson, D. M., 1999. Aging and prospective memory: Differences between naturalistic and laboratory tasks. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 54(4), P256-P269.

Rendell, P. G., Craik, F. I., 2000. Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(7), S43-S62.

Rickard, T. C., Pan, S. C., 2018. A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, 25(3), 847-869.

Ricker, T. J., Vergauwe, E., Cowan, N., 2016. Decay theory of immediate memory: From Brown (1958) to today (2014). *The Quarterly Journal of Experimental Psychology*, 69(10), 1969-1995.

Roediger, H. L., 1973. Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 644-657.

Roediger, H. L., Neely, J. H., 1982. Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 36(2), 213.

Roediger III, H. L., 1996a. Memory illusions. *Journal of memory and Language*, 35(2), 76-100.

Roediger, H. L., Karpicke, J. D., 2006. The Power of Testing Memory. *Perspectives on Psychological Science*, 1(3), 181.

Roediger, H. L. III, Weinstein, Y., Agarwal, P. K., 2010. Forgetting: Preliminary considerations. In S. Della Sala (Ed.), Forgetting (p. 1–22). Psychology Press.

Roediger, H.L., Butler, A.C., 2011. The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science*, 15, 20–27.

Rogers, T. T., McClelland, J. L., 2004. Semantic cognition: A parallel distributed processing approach. MIT press.

Ronis, D.L., Yates, J.F. Kirscht, J.P. (1989). Attitudes, decisions, and habits as determinants of repeated behavior. In A.R. Pratkanis, S.J. Breckler & A.G. Greenwald, (Eds), Attitude Structure and Function. Hillsdale, NJ: Erlbaum, pp. 213-239.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.J. 2017. Bayesian analysis of factorial designs. *Psychological Methods*, 22, 304–321.

Rowland, C. A., 2014. The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463.

Rubin, D. C., Rahhal, T. A., & Poon, L. W. (1998). Things learned in early adulthood are remembered best. *Memory & cognition*, 26(1), 3-19.

Rubin, E. H., Morris, J. C., Grant, E. A., Vendegna, T., 1989. Very mild senile dementia of the Alzheimer type: I. Clinical assessment. *Archives of Neurology*, 46(4), 379-382.

Rubin, D. C., Wenzel, A. E., 1996. One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734– 760.

Ryan, E. B., 1992. Beliefs about memory changes across the adult life span. *Journal of Gerontology*, 47(1), P41-P46.

Rybarczyk, B. D., Hart, R. P., Harkins, S. W., 1987. Age and forgetting rate with pictorial stimuli. *Psychology and Aging*, 2(4), 404.

Saczynski, J. S., Sigurdsson, S., Jonsdottir, M. K., Eiriksdottir, G., Jonsson, P. V., Garcia, M. E., …, Launer, L. J., 2009. Cerebral infarcts and cognitive performance: importance of location and number of infarcts. *Stroke*, 40(3), 677-682.

Sadeh, T., Ozubko, J. D., Winocur, G., Moscovitch, M., 2014. How we forget may depend on how we remember. *Trends in cognitive sciences*, 18(1), 26-36.

Saling, M. M., 2009. Verbal memory in mesial temporal lobe epilepsy: beyond material specificity. *Brain*, 132(3), 570-582.

Salthouse, T. A., 1992. Influence of processing speed on adult age differences in working memory. *Acta psychologica*, 79(2), 155-170.

Saunders, J., Fernandes, M., Kosnes, L., 2009. Retrieval-induced forgetting and mental imagery. *Memory & Cognition*, 37(6), 819-828.

Schacter, D. L., Kihlstrom, J. F., Kaszniak, A. W., Valdiserri, M., 1993. Preserved and impaired memory functions in elderly adults. In J. Cerella, J. M. Rybash, W. Hoyer, & M. L. Commons (Eds.), Adult information processing: Limits on loss (p. 327–350). Academic Press.

Schaffer, G., Poon, L. W., 1982. Individual variability in memory training with the elderly. *Educational Gerontology: An International Quarterly*, 8(3), 217-229.

Schmitter-Edgecombe, M., Woo, E., Greeley, D. R., 2009. Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*, 23(2), 168.

Schnitzspahn, K. M., Kliegel, M., 2009. Age effects in prospective memory performance within older adults: the paradoxical impact of implementation intentions. *European Journal of Ageing*, 6(2) 147–155.

Schnitzspahn, K. M., Zeintl, M., Jäger, T., Kliegel, M., 2011. Metacognition in prospective memory: Are performance predictions accurate?. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(1), 19.

Schnitzspahn, K. M., Kvavilashvili, L., Altgassen, M., 2018. Redefining the pattern of age-prospective memory-paradox: new insights on age effects in lab-based, naturalistic, and self-assigned tasks. *Psychological research*, 1-17.

Schoer, L., 1962. Effect of list length and interpolated learning on the learning and recall of fast and slow learners. *Journal of Educational Psychology*, 53(4), 193-197.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H., 2010. Focal/nonfocal cue effects in PM: monitoring difficulty or different retrieval processes?. Journal of experimental psychology. *Learning, memory, and cognition*, 36(3), 736–749.

Scullin, M., Bugg, J., McDaniel, M., Duberstein, P., 2012. Whoops, I Did It Again: Commission Errors in Prospective Memory. *Psychology and Aging*, 27(1), 46-53.

Scullin, M. K., McDaniel, M. A., Shelton, J. T., 2013. The Dynamic Multiprocess Framework: Evidence from prospective memory with contextual variability. *Cognitive psychology*, 67(1-2), 55-71.

Sekeres, M.J., Winocur, G., & Moscovitch, M., 2018. The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters,* 680, 39–53.

Sheeran, P., Orbell, S.,1999. Implementation intentions and repeated behaviour: Augmenting the predictive validity of the theory of planned behaviour. *European journal of social psychology*, 29(2-3), 349-369.

Shelton, J.T., Lee, J.H., Scullin, M.K., Rose, N.S., Rendell, P.G., McDaniel, M.A., 2016. Improving PM in healthy older adults and individuals with very mild Alzheimer's disease. *Journal of the American Geriatrics Society*, 64(6), 1307–1312.

Shimamura, A. P., 1986. Priming effects in amnesia: Evidence for a dissociable memory function. *The Quarterly Journal of Experimental Psychology*, 38(4), 619-644.

Shuell, T. J., Keppel, G., 1970. Learning ability and retention. *Journal of Educational Psychology*, 61 (1), 59-65.

Shulman, K.I., 2000. Clock-drawing: is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15, 548–561.

Silberzahn, R., et al. 2018. Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science* 1(3), 337-356.

Skaggs, E. B., 1933. A discussion of the temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, 16(3), 411–414.

Slamecka, N. J., 1968. A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin*, 69(6), 423-438.

Slamecka, N. J., McElree, B., 1983. Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(3), 384.

Slamecka, N. J., 1985. On comparing rates of forgetting: Comment on Loftus (1985). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 812-816.

Slamecka, N. J., Katsaiti, L. T., 1988. Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 716-72.

Smallwood, J. M., Baracaia, S. F., Lowe, M., Obonsawin, M., 2003. Task unrelated thought whilst encoding information. *Consciousness and cognition*, 12(3), 452-484.

Smith, G., Del Sala, S., Logie, R., Maylor, E., 2000. Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, 8(5), 311-321.

Smith, R.E., 2003. The cost of remembering to remember in event-based PM: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29. 347–361.

Smith, R.E., Bayen, U.J., 2004. A multinomial model of event-based PM. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30. 756–777.

Smith R. E., Hunt R. R., 2014. Prospective memory in young and older adults: the effects of task importance and ongoing task load. *Aging Neuropsychology and Cognition*, 21, 411–431.

Spitzer, H. F., 1939. Studies in retention. *Journal of Educational Psychology*, 30(9), 641–656.

Spitzer, B., 2014. Finding retrieval-induced forgetting in recognition tests: A case for baseline memory strength. *Frontiers in Psychology*, 5, 1102.

Squire, L. R., 1981. Two forms of human amnesia: An analysis of forgetting. *Journal of Neuroscience,* 1(6), 635-640.

Squire, L. R., Cohen, N. J., Nadel, L., 1984. The medial temporal region and memory consolidation: A new hypothesis. In H. Weingartner & E. S. Parker (Eds.), Memory consolidation (pp. 185-210). Hillsdale, NJ: Erlbaum.

Squire, L. R., 1989. On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 241-245.

Squire, L. R., Alvarez, P., 1995. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, 5(2), 169-177.

Squire, L. R., 2009. Memory and brain systems: 1969–2009. *Journal of Neuroscience*, 29(41), 12711-12716.

Squire, L. R., Bayley, P. J., Smith, C. N., 2009. Amnesia: Declarative and Nondeclarative Memory. *Concise Encyclopedia of Brain and Language*, 30-35.

Squire, L. R., Genzel, L., Wixted, J. T., Morris, R. G., 2015. Memory consolidation. *Cold Spring Harbor perspectives in biology*, 7(8), a021766.

Stamate, A., Logie, R. H., Baddeley, A., Della Sala, S., 2020. Forgetting in Alzheimer's disease: Is it fast? Is it affected by repeated retrieval?. *Neuropsychologia*, 107351.

Storm, B. C., Angello, G., Bjork, E. L., 2011. Thinking can cause forgetting: Memory dynamics in creative problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1287.

Storm, B. C., Levy, B. J., 2012. A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, 40, 827– 843.

Storm, Benjamin C., "Measures of Forgetting", in Handbook of Research Methods in Human Memory ed. Hajime Otani and Bennett L. Schwartz (Abingdon: Routledge, 22 Oct 2018), accessed 29 Apr 2020, Routledge Handbooks Online.

Stroud, J. B., Carter, L. J., 1961. Inhibition phenomena in fast and slow learners. *Journal of Educational Psychology*, 52(1), 30-34.

Tandoh, K., Naka, M., 2007. Durability of retrieval-induced forgetting. *Japanese Journal of Psychology*, 78, 310-315

Taylor, S. E., Schneider, S. K., 1989. Coping and the simulation of events. *Social cognition*, 7(2), 174-194.

Taylor, S. E., Pham, L. B., Rivkin, I. D., Armor, D. A., 1998. Harnessing the imagination: Mental simulation, self-regulation, and coping. *American psychologist*, 53(4), 429.

Thomas, A., Bulevich, J., Zacks, Rose T., 2006. Effective Cue Utilization Reduces Memory Errors in Older Adults. *Psychology and Aging*, 21(2), 379-389.

Thomas, R.C., Weywadt, C.R., Anderson, J.L., Martinez-Papponi, B., McDaniel, M.A., 2018. Testing encourages transfer between factual and application questions in an online learning environment. *Journal of Applied Research in Memory and Cognition*, 7, 252–260.

Thompson, C., Henry, J., Rendell, P., Withall, A., Brodaty, H., 2015. How Valid Are Subjective Ratings of Prospective Memory in Mild Cognitive Impairment and Early Dementia? *Gerontology*, 61(3), 251-257.

Tulving, E., Pearlstone, Z., 1966. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 381-391.

Tulving, E., 1967. The effects of presentation and recall of material in free-recall learning. *Journal of verbal learning and verbal behavior*, 6(2), 175-184.

Tulving, E., 1974. Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist*, 62(1), 74-82.

Tulving E. Elements of episodic memory. Oxford University Press; New York: 1983.

Tulving, E., 1985. How many memory systems are there?. *American psychologist*, 40(4), 385.

Underwood, B. J., 1954. Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, 51, 276–282.

Underwood, B. J., 1964. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 3, 112-129.

Unsworth, N., Brewer, G., Spillers, A., 2011. Variation in working memory capacity and episodic memory: Examining the importance of encoding specificity. *Psychonomic Bulletin & Review*, 18(6), 1113–1118.

Uttl, B., Graf, P., Richter, L.K., 2002. Verbal Paired Associate tests: Limits on validity and reliability. *Archives of Clinical Neuropsychology*, 17, 567–581.

Uttl, B., 2005. Measurement of Individual Differences: Lessons From Memory Assessment in Research and Clinical Practice. *Psychological Science*, 16(6), 460–467.

Uttl, B., 2006. Age-related changes in event-cued visual and auditory prospective memory proper. *Aging, Neuropsychology, and Cognition*, 13(2), 141-172.

Uttl, B., 2008. Transparent meta-analysis of prospective memory and aging. *PloS one*, 3(2).

Vakil, E., 2005. The Effect of Moderate to Severe Traumatic Brain Injury (TBI) on Different Aspects of Memory: A Selective Review. *Journal of Clinical and Experimental Neuropsychology*, 27(8), 977–1021.

Vallet, G., Rouleau, I., Benoit, S., Langlois, R., Barbeau, E., Joubert, S., 2016. Alzheimer's disease and memory strength: gradual decline of memory traces as a function of their strength. *Journal of Clinical and Experimental Neuropsychology*, 38(6), 648–660.

Verde, M. F., 2012. Retrieval-induced forgetting and inhibition: A critical review. In B. H. Ross (Ed.), *Psychology of learning and motivation*, 56, 47–80.

Verhaegen, P., Martin, M., Sedek, G., 2012. Reconnecting cognition in the lab and cognition in real life: The role of compensatory social and motivational factors in explaining how cognition ages in the wild. *Aging, Neuropsychology and Cognition*, 19, 1–12.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., Kievit, R. A., 2012. An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632-638.

Walker, M. P., Brakefield, T., Hobson, J. A., Stickgold, R., 2003. Dissociable stages of human memory consolidation and reconsolidation. *Nature*, 425(6958), 616-620.

Walser, M., Fischer, R., Goschke, T., 2012. The failure of deactivating intentions: Aftereffects of completed intentions in the repeated prospective memory cue paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 1030.

Walser, M., Goschke, T., Möschl, M., Fischer, R., 2017. Intention deactivation: Effects of prospective memory task similarity on aftereffects of completed intentions. *Psychological Research*, 81(5), 961-981.

Walsh, C.M., Wilkins, S., Bettcher, B.M., Butler, C.R., Miller, B.L., Kramer, J.H., 2014. Memory consolidation in aging and MCI after 1 week. *Neuropsychology*, 28, 273–280.

Walter, S., Meier, B., 2014. How important is importance for PM? A review. *Frontiers in Psychology*, 5, 657.

Weingartner, H., Parker, E.S. (eds) 1984. Memory consolidation : psychobiology of cognition Lawrence Erlbaum: Hillsdale, New Jersey

Welsh K, Butters N, Hughes J, Mohs R, Heyman A., 1991. Detection of Abnormal Memory Decline in Mild Cases of Alzheimer's Disease Using CERAD Neuropsychological Measures. *Archives of neurology*, 48(3):278–281.

West, R., 2011. The temporal dynamics of prospective memory: A review of the ERP and prospective memory literature. *Neuropsychologia*, 49(8), 2233-2245.

Weston, P.S.J., Nicholas, J.M., Henley, S.M.D., Liang, Y., Macpherson, K., et al., 2018. Accelerated long-term forgetting in presymptomatic autosomal dominant Alzheimer's disease: a cross-sectional study. *Lancet Neurology*, 17(2), 123–132.

Wheeler, M. A., Roediger, H. L., III, 1992. Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240-245.

Wheeler, M. A., Ewers, M., Buonanno, J. F., 2003. Different rates of forgetting following study versus test trials. *Memory*, 11, 571–580.

Wieber, F., von Suchodoletz, A., Heikamp, T., Trommsdorff, G., Gollwitzer, P. M., 2011. If-then planning helps school-aged children to ignore attractive distractions. *Social Psychology*, 42(1), 39–47.

Wilkins, D., 1976. Notional syllabuses. Bulletin CILA (Commission interuniversitaire suisse de linguistique appliquée)(«Bulletin VALS-ASLA» depuis 1994), 24, 5-17.

Wilkins, A.J., Baddeley, A.D., 1978. Remembering to recall in everyday life: An approach to absent-mindedness. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory. New York: Academic Press.

Wilkinson, H., Holdstock, J. S., Baker, G., Herbert, A., Clague, F., Downes, J. J., 2012. Long-term accelerated forgetting of verbal and non-verbal information in temporal lobe epilepsy. *Cortex*, 48(3), 317-332.

Wilson, R.S., Segawa, E., Boyle, P.A., Anagnos, S.E., Hizel, L.P., Bennett, D.A., 2012. The natural history of cognitive decline in Alzheimer's disease. *Psychology and Aging*, 27, 1008–17.

Wimer, R. E., 1960. Age differences in incidental and intentional learning. *Journal of Gerontology*, 15, 79–82.

Winter, M., 1919. Aesops fables for children with pictures. United States, Washington: Rand, McNally & Co. Retrieved from http://read.gov/aesop/about.html

Wirth, M., Bäuml, K.T., 2020. Category labels can influence the effects of selective retrieval on nonretrieved items. *Memory & Cognition*, 48**,** 481–493.

Wixted, J. T. 1990., Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 927.

Wixted, J. T., Ebbesen, E. B., 1997. Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, 25, 731-739.

Wixted, J. T., 2004. The psychology and neuroscience of forgetting. *The Annual Review of Psychology*, 55, 235-269.

Wixted, J. T., 2007. Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, 114(1), 152.

Woods, S. P., Iudicello, J. E., Moran, L. M., Carey, C. L., Dawson, M. S., & Grant, I., 2008. HIV-associated prospective memory impairment increases risk of dependence in everyday functioning. *Neuropsychology*, 22(1), 110.

Yan, J.H., Dick, M.B., 2006. Practice effects on motor control in healthy seniors and patients with mild cognitive impairment and Alzheimer’s disease. *Neuropsychology, Development, and Cognition Section B, Aging, Neuropsychology and Cognition*, 13 (3–4), 385–410.

Yang, J. et al., 2016. Effects of learning experience on forgetting rates of item and associative memories. *Learning & memory* (Cold Spring Harbor, N.Y.), 23(7), 365–378.

Yano, M., Umeda, S., Mimura, M., 2008. Preserved priming but insensitivity to perceptual fluency on recognition judgments in Alzheimer's disease. *Psychogeriatrics*, 8(4), 178-187.

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M.B., Leirer, V.O., 1983. Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, 17, 37–49.

Zacks, R. T., Hasher, L., Li, K. Z., 2000. Human memory. F.I.M Second Edition, T.A Craik and, Salthouse (Eds.), The Handbook of Aging and Cognition, Lawrence Erlbaum Associates, Mahwah, NJ (2000), 2093-3057.

Zeman, A. Z., Boniface, S. J., Hodges, J. R., 1998. Transient epileptic amnesia: a description of the clinical and neuropsychological features in 10 cases and a review of the literature. *Journal of Neurology, Neurosurgery & Psychiatry*, 64(4), 435-443.

Zerr, C.L. 2017. The domain-generality and durability of efficient learning (Master's thesis, Washington University in St. Louis). Retrieved from: https://doi.org/10.7936/K7MP52PD

Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., McDermott, K. B. 2018. Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. *Psychological Science*, 29, 1436–1450.

Zimmermann, T. D., Meier, B., 2010. The effect of implementation intentions on prospective memory performance across the lifespan. *Applied Cognitive Psychology*, 24(5), 645-658.

APPENDICES

1. Supplementary material for Experiment 1 and 2.

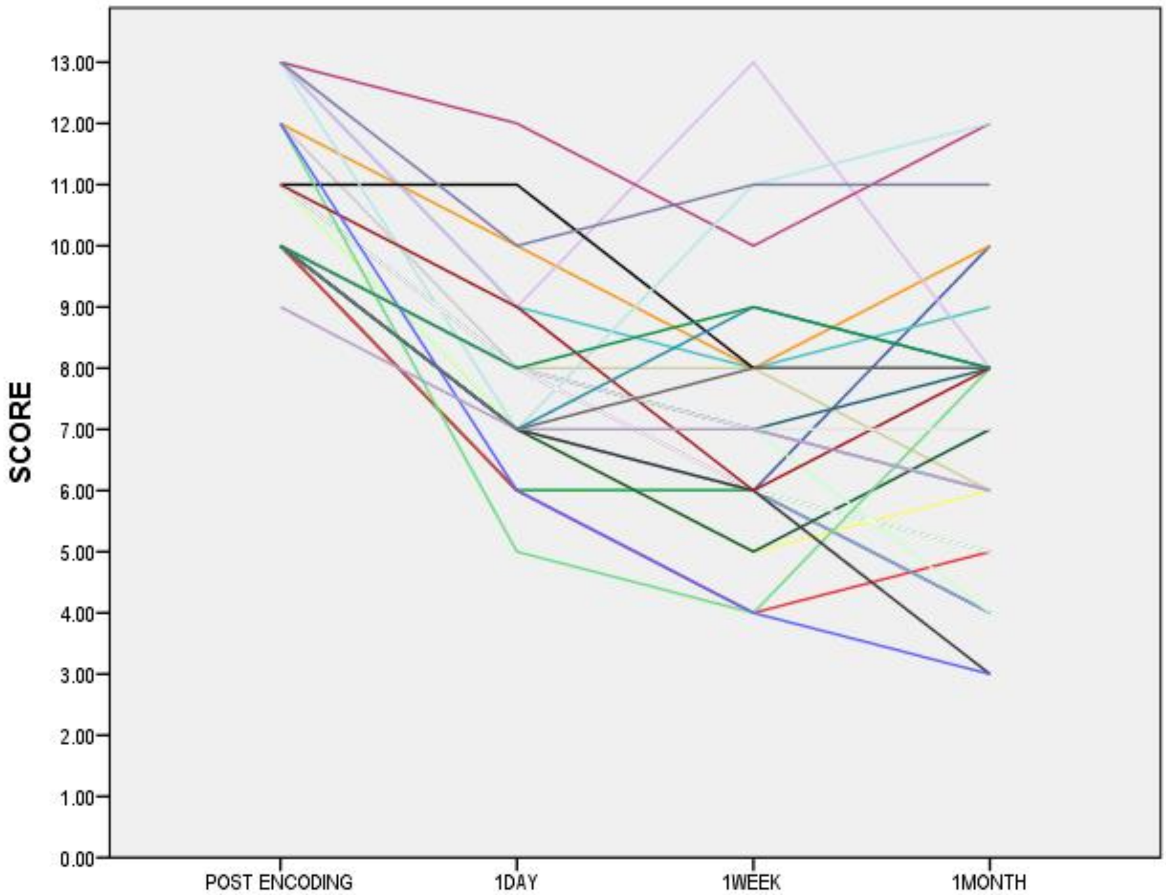


Figure 2.1a. Individual recall performance of the younger participants at different delay intervals in the condition with retrieval practice with integrated material.

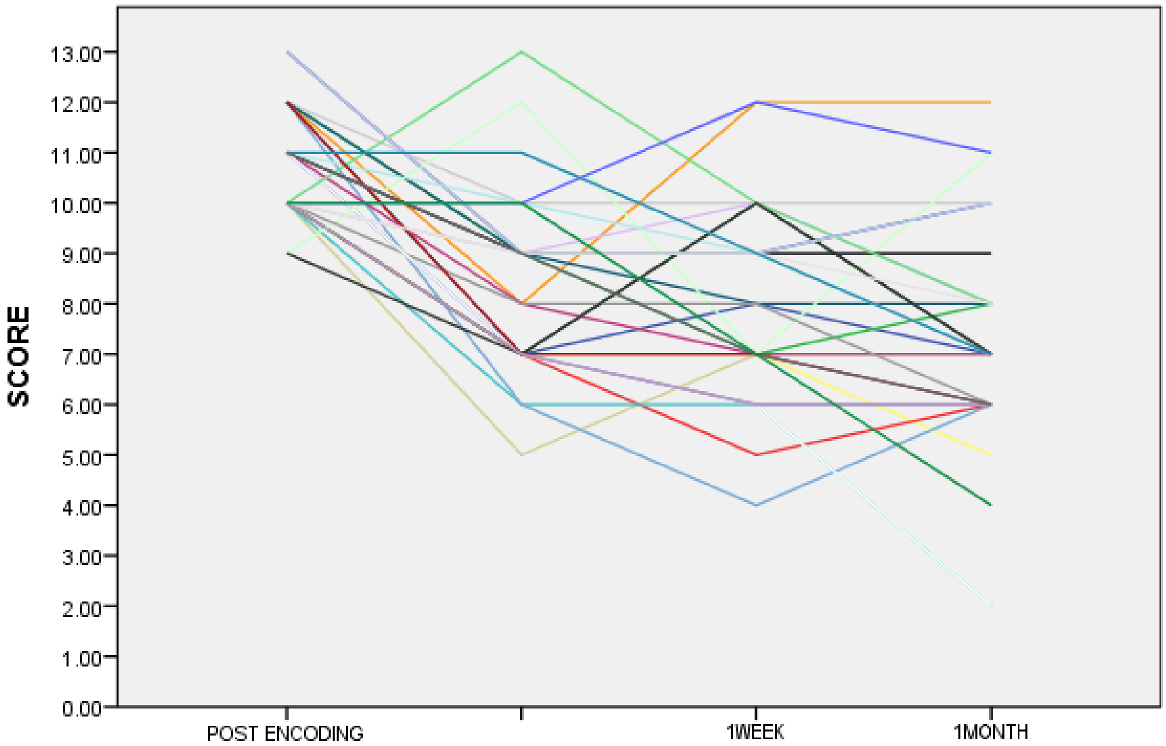


Figure 2.1b. Individual recall performance of the older participants at different delay intervals in the condition retrieval practice with integrated material.

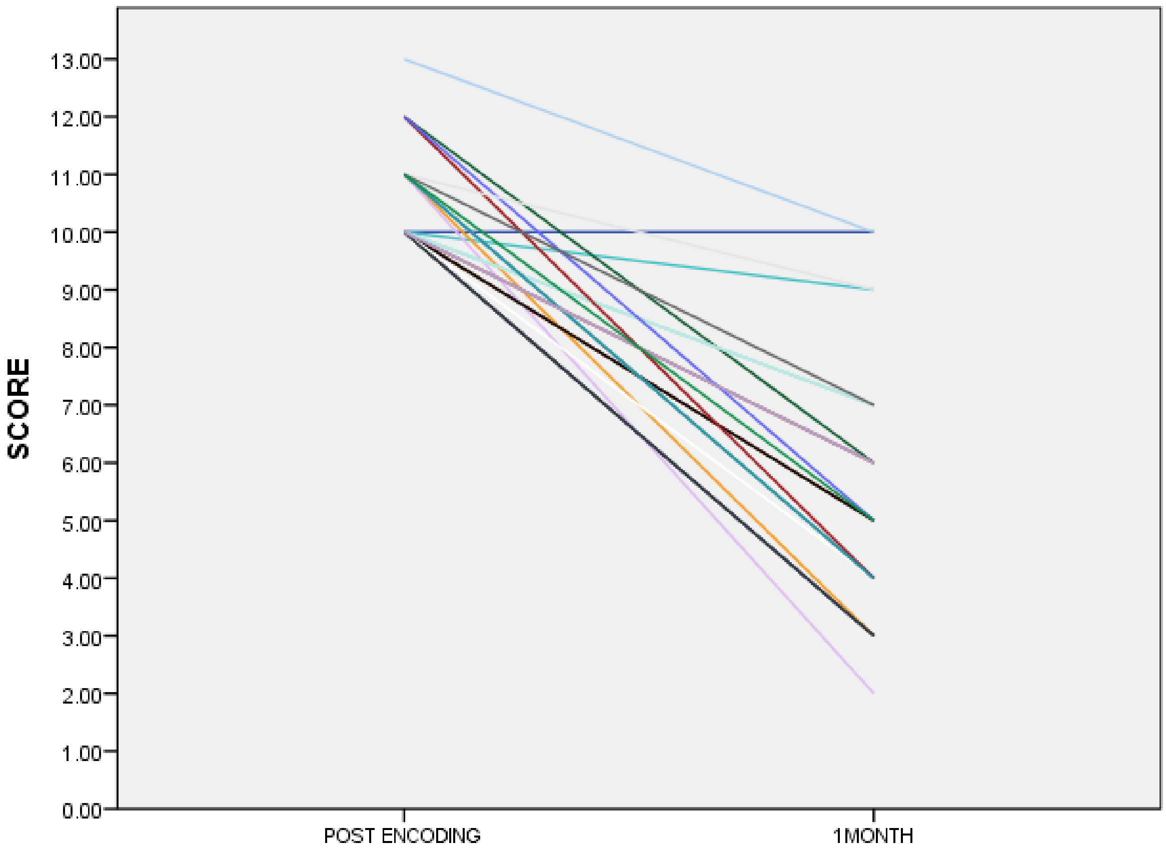


Figure 2.1c. Individual recall performance of the younger participants at different delay intervals in the condition without retrieval practice with integrated material.

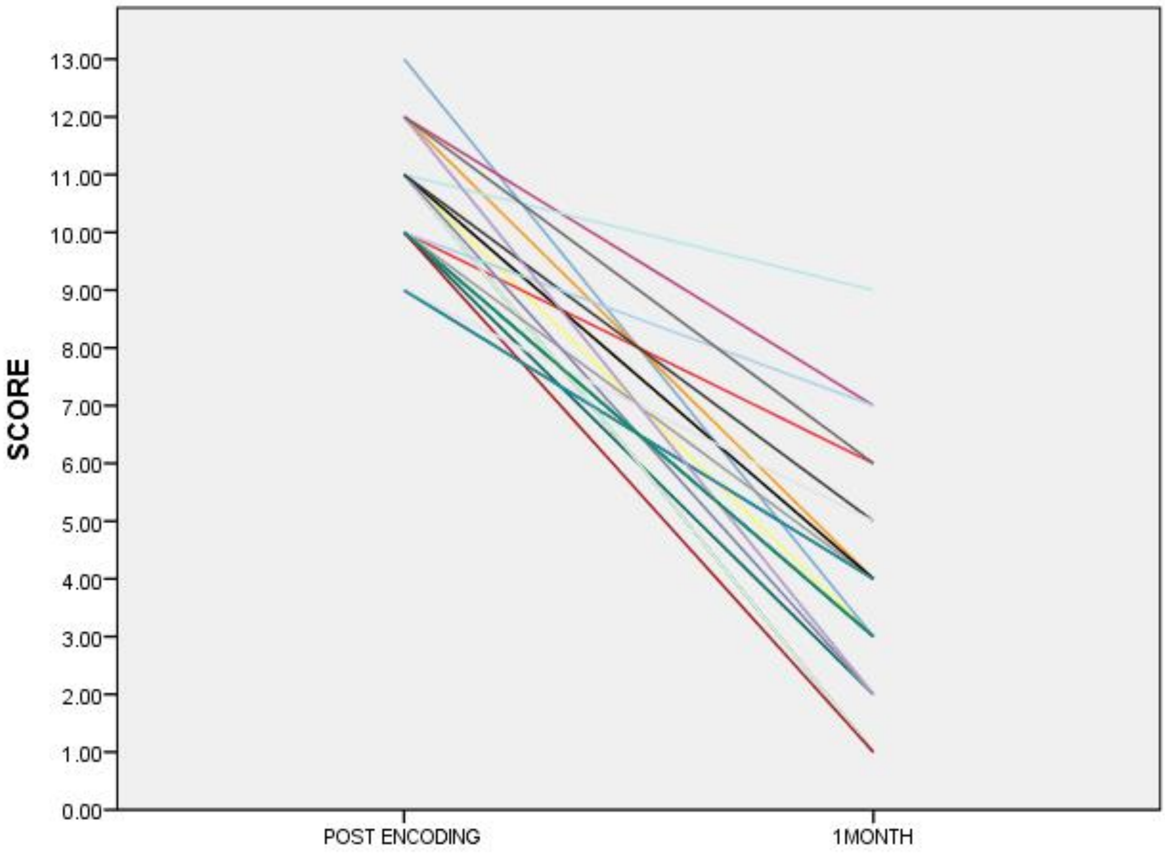


Figure 2.1d. Individual recall performance of the older participants at different delay intervals in the condition without retrieval practice with integrated material.

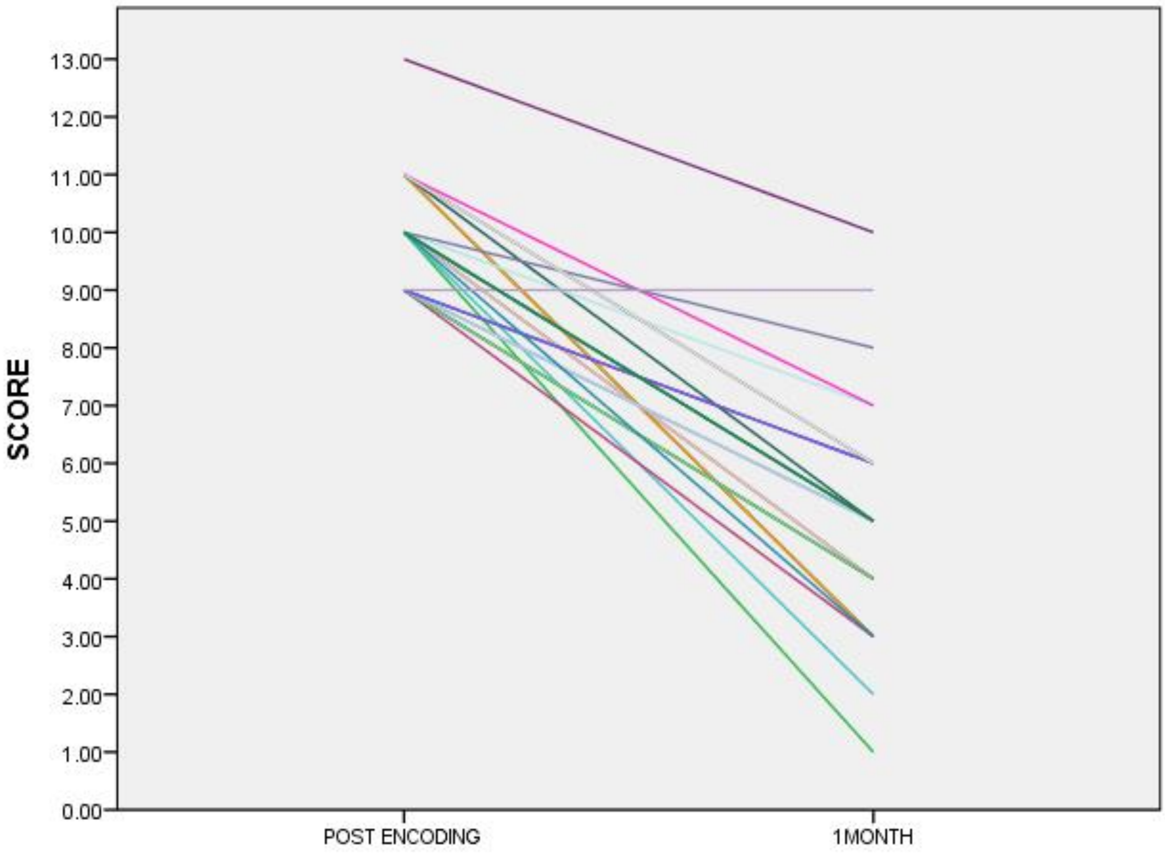


Figure 2.2a. Individual recall performance of younger participants at different delay intervals in the condition without retrieval practice on material with the disrupted narrative.

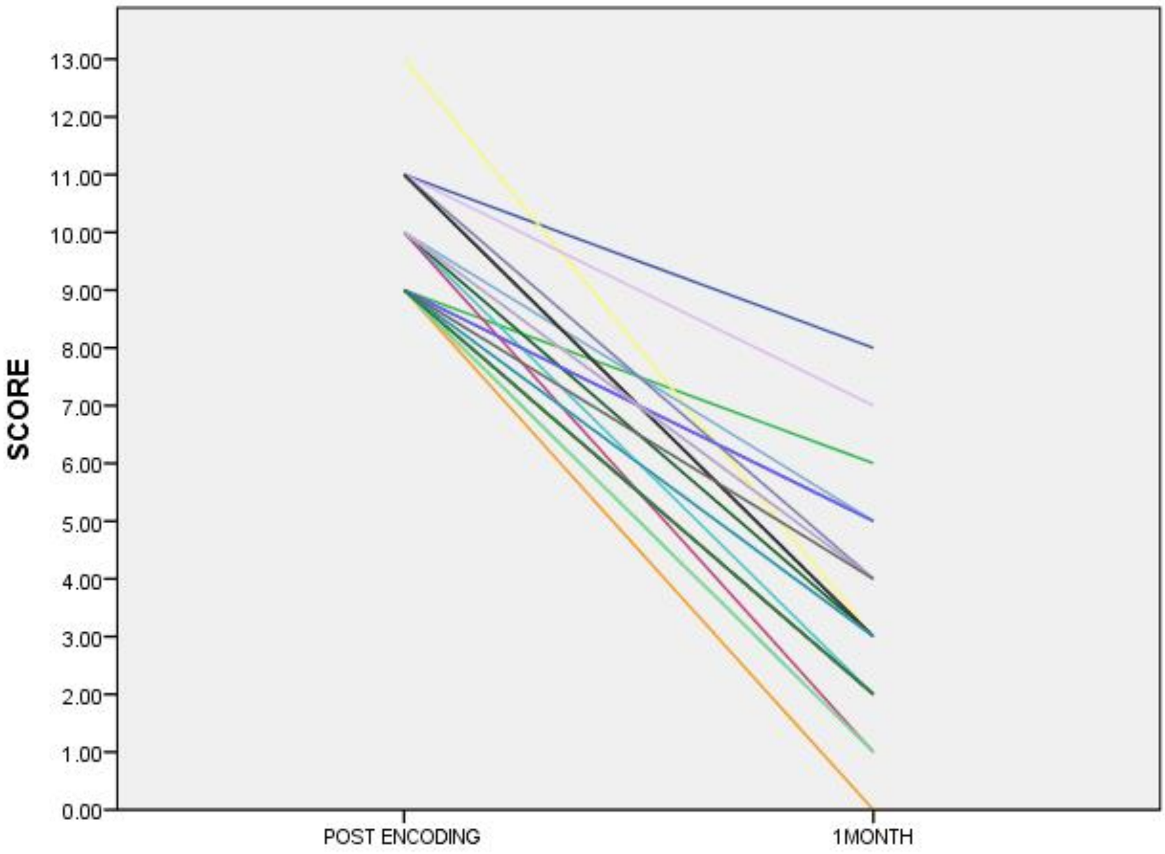


Figure 2. 2b. Individual recall performance of the older participants at different delay intervals the condition without retrieval practice on material with the disrupted narrative.

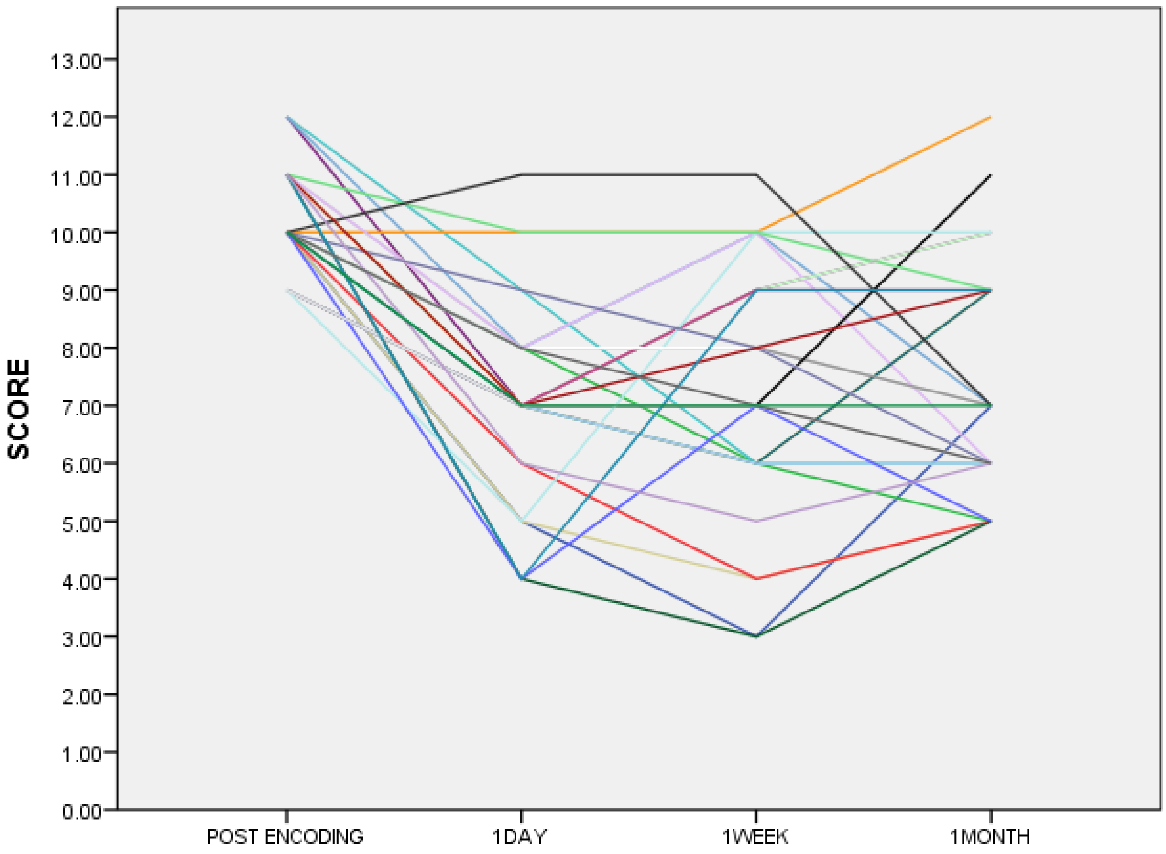


Figure 2.2c. Individual recall performance of the younger participants at different delay intervals in the condition with retrieval practice on material with the disrupted narrative.

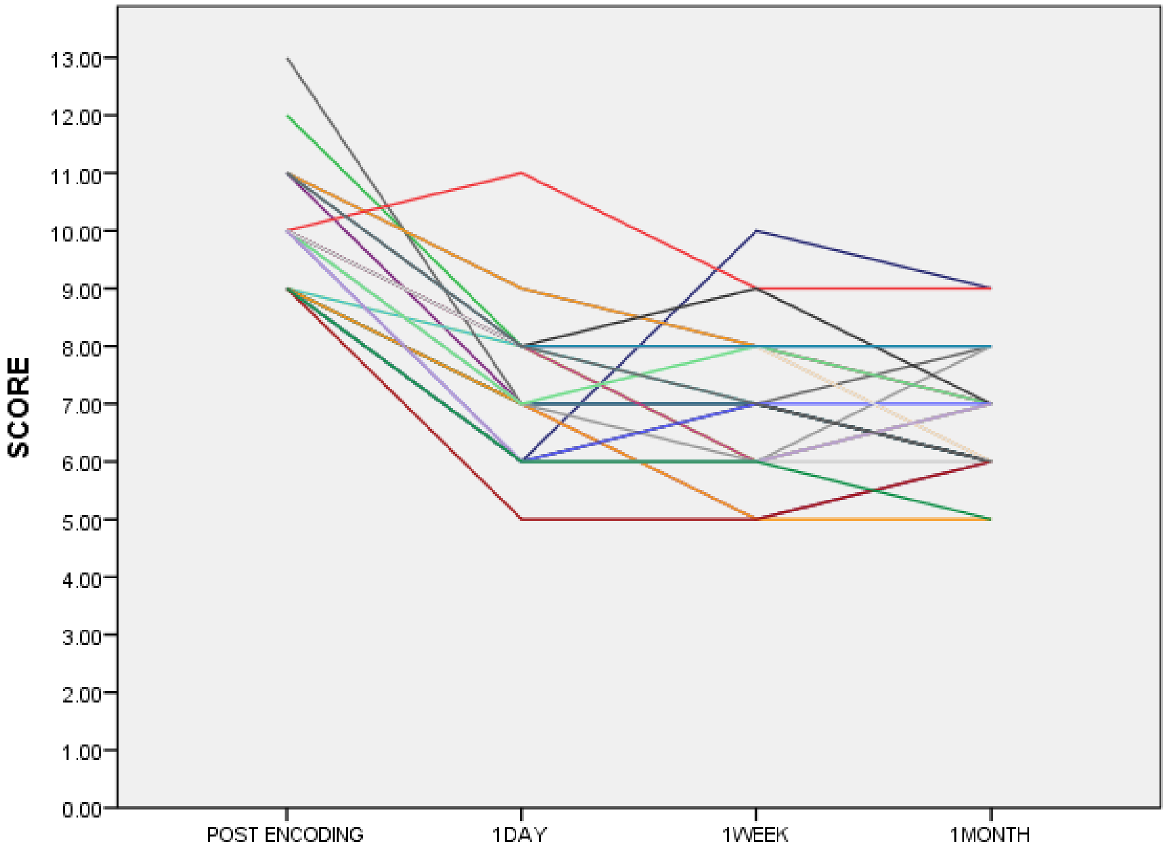


Figure 2.2d. Individual recall performance of the older participants at different delay intervals in the condition with retrieval practice on material with the disrupted narrative.

Participant instructions:

“You will be read four short stories. Each story has several common features which you should remember: age, sex, nationality, characteristic of the main character, the animal or animals, characteristic of the animal, the time, place, action, and moral of each story. We realise this is a rather difficult exercise; therefore, you probably won't remember all of it, but do your best.

Any questions?

Then here are the stories:

(when reading, pause for 2 sec after each sentence and for 5 s between stories) “

Fables - Integrated Material:

Story 1 - Reacting in anger.

A French woman used to take long walks up on a hill and came across a swarm of bees.

One cloudy afternoon the young woman approached the swarm when two bees flew out and stung her.

She flew into an angry rage and started swinging for the bees with a stick, but a lot of tiny bees flew out and attacked her.

The moral of the story with the tiny bees is about reacting in anger: If you react in anger you can make the situation worst.

Story 2 - Good deeds.

A Spanish woman lived quietly in her white house in a small village with a lot of dogs.

One sunny morning the elderly woman saw a group of dogs fighting a frightened small dog, so she went out and separated them.

The compassionate woman gained a friend by saving the small dog, which remained by her side and protected her for many years after.

The moral of the story with the frightened small dog is about good deeds: One good deed deserves another.

Story 3 – Pleasure.

A Korean man used to play around a pond that was usually full of frogs swimming about in the water.

One starry evening the young man started to throw rocks at the frogs and counted how many he succeeded in hitting.

The man was very proud of his aiming skills, but a wise frog appeared in his dream and advised him to stop throwing rocks at frogs.
The moral of the story with the wise frog is about pleasure: We should not take pleasure at the expense of others.

Story 4 – Stubbornness.
An Indian man living at the foot of a mountain used to lead his donkey down a road to the nearest market.
One Wednesday the donkey got free and ran towards the edge of a high cliff, so the elderly man ran to catch him.
The man was worried that the donkey might fall so he grabbed its tail, but the strong-willed animal escaped and fell off.
The moral of the story with the strong-willed donkey is about stubbornness: Stubbornness can sometimes lead to perdition.

Sentences - Material with the disrupted narrative:
There were three versions, with the sentences scrambled in a different way on each, to allow for multiple encoding trials.

Version 1.
The moral of the story with the wise frog is about pleasure: We should not take pleasure at the expense of others.
A Spanish woman lived quietly in her white house in a small village with a lot of dogs.
The man was worried that the donkey might fall so he grabbed its tail, but the strong-willed animal escaped and fell off.
One cloudy afternoon the young woman approached the swarm when two bees flew out and stung her.

A Korean man used to play around a pond that was usually full of frogs swimming about in the water.
The compassionate woman gained a friend by saving the small dog, which remained by her side and protected her for many years after.

One Wednesday the donkey got free and ran towards the edge of a high cliff, so the elderly man ran to catch him.
The moral of the story with the tiny bees is about reacting in anger: by reacting in anger one may make things worse.

An Indian man living at the foot of a mountain used to lead his donkey down a road to the nearest market.
One sunny morning the elderly woman saw a group of dogs fighting a frightened small dog, so she went out and separated them.
The man was very proud of his aiming skills, but a wise frog appeared in his dream and advised him to stop throwing rocks at frogs.
She flew into an angry rage and started swinging for the bees with a stick, but a lot of tiny bees flew out and attacked her.

The moral of the story with the strong-willed donkey is about stubbornness.
A French woman used to take long walks up on a hill and came across a swarm of bees.
One starry evening the young man started to throw rocks at the frogs and counted how many he succeeded in hitting.
The moral of the story with the frightened small dog is about good deeds: One good deed deserves another.

Question Sets (1-4):
The same sets were used for both types of material (Integrated Material; Material with the disrupted narrative).

Set 1.
1. What was the sex of the French person? Woman
2. What animal was saved in a story? Dog
3. What was the age sex of the person in the story that took place in the evening? Young man
4. What animal/animals were involved in the story with a French person? Bees
5. What was the sex of the person who was compassionate? Woman
6. What animal/animals were involved in the story with a young man? Frogs
7. What animal/animals were involved in the story with a compassionate person? Dog

8. What was the sex of the person in the story that took place on a hill? Woman
9. What is the moral of the story with the donkey about? Stubbornness
10. What animal/animals were involved in the story that took place in the evening? Frogs
11. What was the sex of the person who made a friend? Woman
12. What was the nationality of the person in the story that took place on a hill? French
13. How was the donkey in the story? Strong willed

Set 2.

1. What was the sex of the person from Spain? Woman
2. What animals were sung at with a stick? Bees
3. What is the moral of the story with frogs about? Pleasures
4. What was the age sex of the person in the story that took place on a Wednesday? Elderly man
5. How was the frog in the story? Wise
6. What animal/animals were involved in the story with an elderly man? Donkey
7. What was the sex of the person in the story that took place in a village? Woman
8. What animal/animals were involved in the story with an angry person? Bees
9. What animal/animals were involved in the story that took place on Wednesday? Donkey
10. What was the nationality of the person in the story that took place in a village? Spanish
11. What was the sex of the person who got angry? Woman
12. What animal/animals were involved in the story with a Spanish person? Dogs
13. What was the sex of the person who was stung? Woman

Set 3.

1. When did the story about animals fighting happen? Morning
2. What was the nationality of the person in the story that took place at the foot of a mountain? Indian
3. What is the moral of the story with bees about? Reacting in anger
4. What was the age/sex of the person in the story that took place in the morning? Elderly woman
5. What was the sex of person from India? Man
6. What animal/animals were involved in the story that took place in the morning? Dogs
7. What was the sex of the person who was proud? Man
8. How (size) were the bees in the story? Tiny

9. What animal/animals were involved in the story with a proud person? Frog/s
10. What was the sex of the person in the story that took place at the foot of a mountain? Man
11. What animal/animals was/were hit with stones in a story? Frog/s
12. What animal/animals were involved in the story with an Indian person? Donkey
13. What was the sex of the person who was advised by an animal? Man

Set 4.

1. What was the age/sex of the person in the story that took place in the afternoon? Young woman
2. What animal fell off a cliff in a story? Donkey
3. What was the sex of the person in the story that took place near a pond? Man
4. What animal/animals were involved in the story with a worried person? Donkey
5. How was the dog in the story? Small
6. What was the sex of person from Korea? Man
7. What animal animal/animals were involved in the story with a young woman? Bees
8. Where did the story about a Korean person happen? Pond
9. What was the sex of the person who was worried? Man
10. What animal/animals were involved in the story that took place in the afternoon? Bees
11. What is the moral of the story with a dog about? Good deeds
12. What animal/animals were involved in the story with a Korean person? Frog/s
13. What was the sex of the person who tried to grab an animal? Man

Probing Scheme.

Story 1

A French woman used to take long walks up on a hill and came across a swarm of bees.
One cloudy afternoon the young woman approached the swarm when two bees flew out and stung her.
She flew into an angry rage and started swinging for the bees with a stick, but a lot of tiny bees flew out and attacked her.
The moral of the story with the tiny bees is about reacting in anger: If you react in anger you can make the situation worst.

Story 2

A Spanish woman lived quietly in her white house in a small village with a lot of dogs.
One sunny morning the elderly woman saw a group of dogs fighting a frightened small dog, so she went out and separated them.
The compassionate woman gained a friend by saving the small dog, which remained by her side and protected her for many years after.
The moral of the story with the frightened small dog is about good deeds: One good deed deserves another.

Set 1 -
Questions 1,4,8,12 probe Story 1/ Sentence 1
Questions 2,5,7,11 probe Story 2/ Sentence 3
Questions 3,6,10 probe Story 3/ Sentence 2
Questions 9 & 13 probe Story 4/ Sentence 4

Story 1/ Sentence 1
- **S1Q1 -** What was the sex of the French person? Woman
- **S1 Q4 -** What animal/animals were involved in the story with a French person? Bees
- **S1 Q8 -** What was the sex of the person in the story that took place on a hill? Woman
- **S1 Q12 -** What was the nationality of the person in the story that took place on a hill? French

Story 2 / Sentence 3
- **S1Q2 -** What animal was saved in a story? Dog
- **S1Q5 -** What was the sex of the person who was compassionate? Woman
- **S1Q7 -** What animal/animals were involved in the story with a compassionate person? Dog
- **S1Q11 -** What was the sex of the person who made a friend? Woman

2. Supplementary material for Experiment 3

Material:

The initial Fables material was build based on The Crimes test but we increased the number of features (13 in our case as opposed to just 5 in the crimes test) so we could do partial testing without ever testing the same feature twice (e.g., moral, nationality, sex, etc.). This was too big of a load (features) for patients; I therefore decided to try several changes. The first thing I did was to make the sentences a bit shorter and reduce the number of features. I wanted to see if it would be easier if I reduced the number of stories presented from 4 to 3. I therefore needed to create a new set of questions.

Pilot 1: Removing one story (present 3 rather than 4).

Removing one story did not seem to work because many features within a set of questions had to be repeated; therefore, when patients did not remember a feature, they automatically lost even more points within that set.

Pilot 2: I tried a 4 story version, with a reduced number of features (8 as opposed to 13). Because 5 features had to be removed, I still needed to create 4 new sets of questions. In addition, by doing this, the features had to be repeated twice in different sets by using reversed questions (e.g., “What was the nationality of the person from the story that took place on a mountain?” Answer: Indian; and “Where did the story with the Indian man take place” Answer: mountain).

Participant instructions:

“You will be read four short stories. Each story has eight common features which you should remember: the age, sex and nationality of the main character, the animal or animals, the time, place action, and moral of each story. We realise this is a rather difficult exercise; therefore, you probably won't remember all of it, but do your best.

Any questions?

Then here are the stories:

(when reading, pause for 2 sec after each sentence and for 5 s between stories) ”

Final fables material:

Story 1 - Reacting in anger.

A young French woman used to take walks through a vineyard.

At the weekend, on a Saturday, the woman found a swarm of bees and two bees flew out and stung her.

She flew into an angry rage and started swinging for the bees with a stick, but a lot of tiny bees flew out and covered her in stings.

The moral of the story with the tiny bees is about reacting in anger: If you react in anger you can make the situation worst.

Story 2 - Good deeds.

An old Spanish woman lived in a small village with a lot of dogs.

One sunny morning, at the break of dawn, the woman saw a small dog with a broken leg, so she went out and cared for him.

The woman gained a friend - the small dog remained by her side and protected her for many years after.

The moral of the story with the frightened small dog is about good deeds: One good deed deserves another.

Story 3 – Pleasure.

A young Italian man used to play around a pond that was usually full of frogs.

One evening, at dusk the man started to throw rocks at the frogs.

But a wise frog soon appeared in one of his dreams and advised him to stop taking pleasure in hitting frogs.

The moral of the story with the frog is about pleasure: We should not take pleasure at the expense of others.

Story 4 – Stubbornness.

An old Indian man lived on a mountain with his donkey.

In the middle of one week, on a Wednesday, the donkey got free and started running towards the edge of cliff.

The man ran to catch him and grabbed its tail, but the animal kicked him, escaped and fell off.

The moral of the story with the strong-willed donkey is about stubbornness: stubbornness can sometimes lead to perdition

Question Sets (1-4):

*St: set; S: sentence.

Set 1:

1. What was the sex of the French person? woman (St1 S1)
2. What animal/animals were involved in the story that took place near a pond? frogs (St 3 S1)
3. What was the nationality of the old man? Indian (St4 S1)
4. What was the nationality of the person in the story that took place near a pond? Italian (St3 S1 -reversed question)
5. What was the sex of the person in the story that took place on a Wednesday? Man (St4 S2)
6. What was the sex and age of the person in the story with the dog? Old woman (St2 S1)
7. What animal/animals were involved in the story that took place on a Wednesday? Donkey (St4 S2)
8. What animal/animals were sung at with a stick? Bees (St1 S3)
9. What was the nationality of the person who grabbed an animal by its tail? Indian (St4 S3)
10. What animal/animals remained by someone's side for many years? Small dog (St2 S3)
11. What is the moral of the story with the person from India about? Stubbornness (St4 S4
12. What is the moral of the story with the dog? Good deeds (St2 S4)
13. What animal/animals were involved in the story about stubbornness? Donkey (St4 S4)

Set 2:

1. What was the sex of the person from India? Man (St4 S1)
2. What animal/animals were involved in the story that took place in a vineyard? Bees (St1 S1)
3. What was the nationality of the old woman? Spain (St2 S1)
4. What was the nationality of the person in the story which took place in a vineyard? French (St1 S1 - reversed question)

5. What was the sex of the person in the story that took place in the morning? (St2 S2)
6. What was the sex and age of the person in the story with the frogs? Young Man (St4 S1)
7. What animal/animals were involved in the story that took place in the morning? (St2 S2)
8. What animal/animals was/were hit with stones in a story? Frogs (St3 S3)
9. What was the nationality of the person who gained a friend? Spanish (St2 S3)
10. What animal/animals appeared in a person's dream? Frog (St3 S3)
11. What is the moral of the story with the person from Spain? Good deeds (St2 S4)
12. What is the moral of the story with the donkey? Stubbornness (St4 S4)
13. What animal/animals were involved in the story about good deeds? Dogs (St2 S4)

Set 3:

1. What was the sex of the person from Spain? Woman (St2 S1)
2. What animal/animals were involved in the story that took place on a mountain? Donkey (St4, S1)
3. What was the nationality of the young man? Spain (St3 S1)
4. What was the nationality of the person from the story that took place on a mountain? Indian (St4, S1)
5. What was the sex of the person in the story that took place at night? Man (St3 S1)
6. What was the sex and age of the person in the story with the bees? Young Woman (St1 S1)
7. What animal/animals were involved in the story that took place at night? Frogs (St3 S2)
8. What animal/animals were cared for in a story? Small dog (St2 S2)
9. What was the nationality of the person who saw an animal in its dream? Italian (St3 S3)
10. What animal/animals stung a person? Bees (St1 S3)
11. What is the moral of the story with the person from Italy? Pleasure (St3 S4)
12. What is the moral of the story with the bees? Reacting in Anger (St1 S4)
13. What animal/animals were involved in the story about pleasure? Frogs (St3 S4)

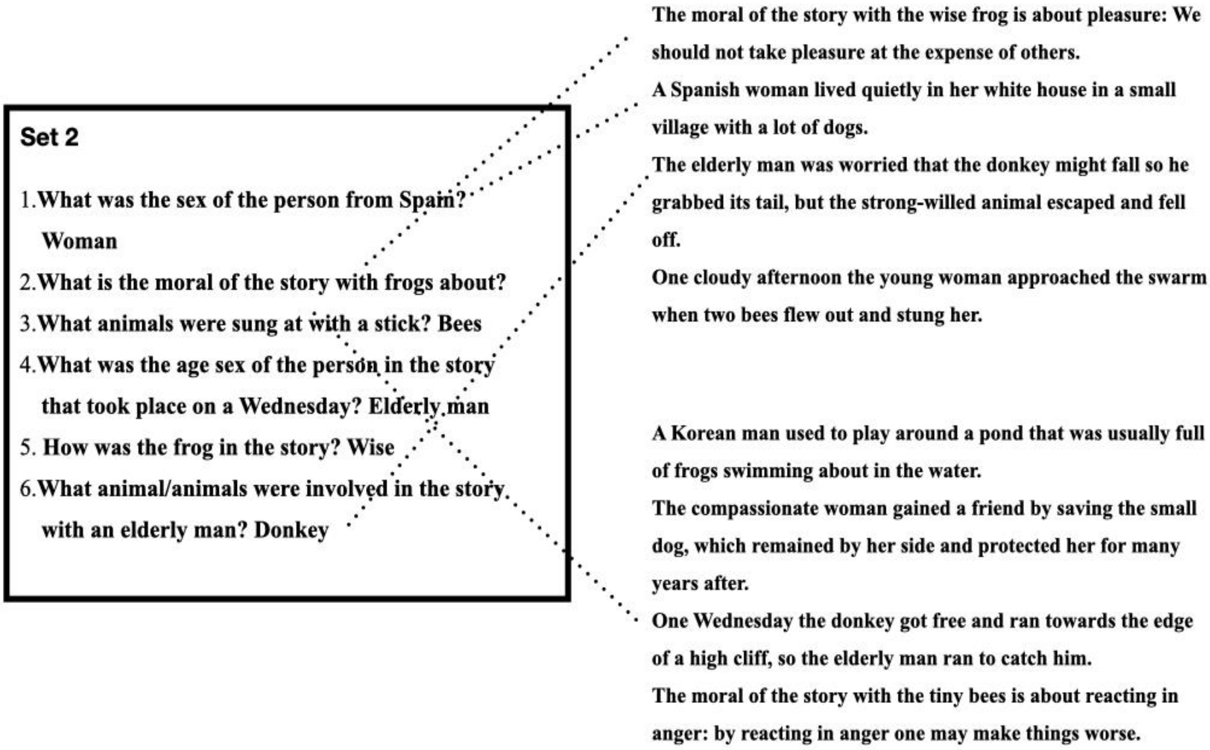
Set 4:

1. What was the sex of the person from Italy? Man (St3 S1)
2. What animal/animals were involved in the story that took place in a village? Dogs (St2 S1)

3. What was the nationality of the young woman? French (St1 S1)
4. What was the nationality of the person in the story that took place in a village? Spanish (St2 S1)
5. What was the nationality of the person in the story that took place on a Saturday? French (St1 S2)
6. What was the sex and age of the person in the story with the donkey? Old man (St4 S1)
7. What animal/animals was/were involved in the story that took place on a Saturday? Bees (St1 S2)
8. What animal/animals fell off a cliff? Donkey (St4 S3)
9. What was the nationality of the person which flew into an angry rage? French (St1 S3)
10. What animal/animals kicked a person? Donkey (St4 S3)
11. What is the moral of the story with the person from France? Reacting in Anger (St1 S4)
12. What is the moral of the story with the frogs? Pleasures (St3 S4)
13. What animal/animals were involved in the story about reacting in anger? Bees (St1 S4)

Probing Scheme:

An example of how questions within a set (e.g. set 2) probe different sentences.



Individual performance data and tables Experiment 3.

Analysis of the individual forgetting curves across all AD patients and HC participants in the condition with retrieval practice revealed considerable individual variability (Fig. 1). In terms of individual scores 3 out of 21 AD participants showed no decline between immediate and one day (1 participant improved); 6 participants showed stable or improved performance between one day and one week testing (5 improved); 6 participants showed no decline between one week and one month; 9 participants had a relatively stable performance across 1 day and 1 month interval.

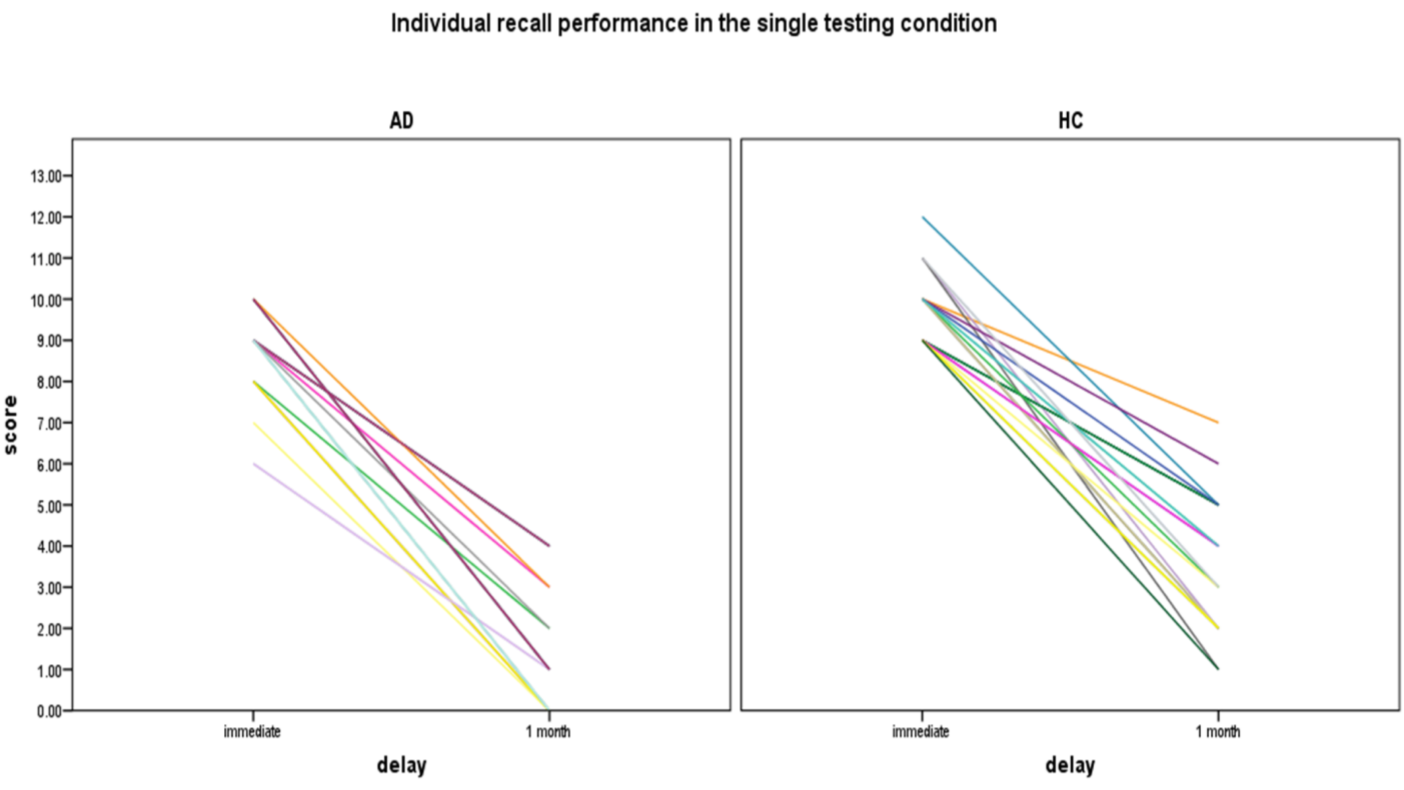


Fig. 3.1. Individual recall performance on the Fables test in the immediate and 1-month tests as a function of test session in the AD and HC groups.
AD: Alzheimer's disease; HC: Healthy controls.

3. Supplementary material for Experiment 5

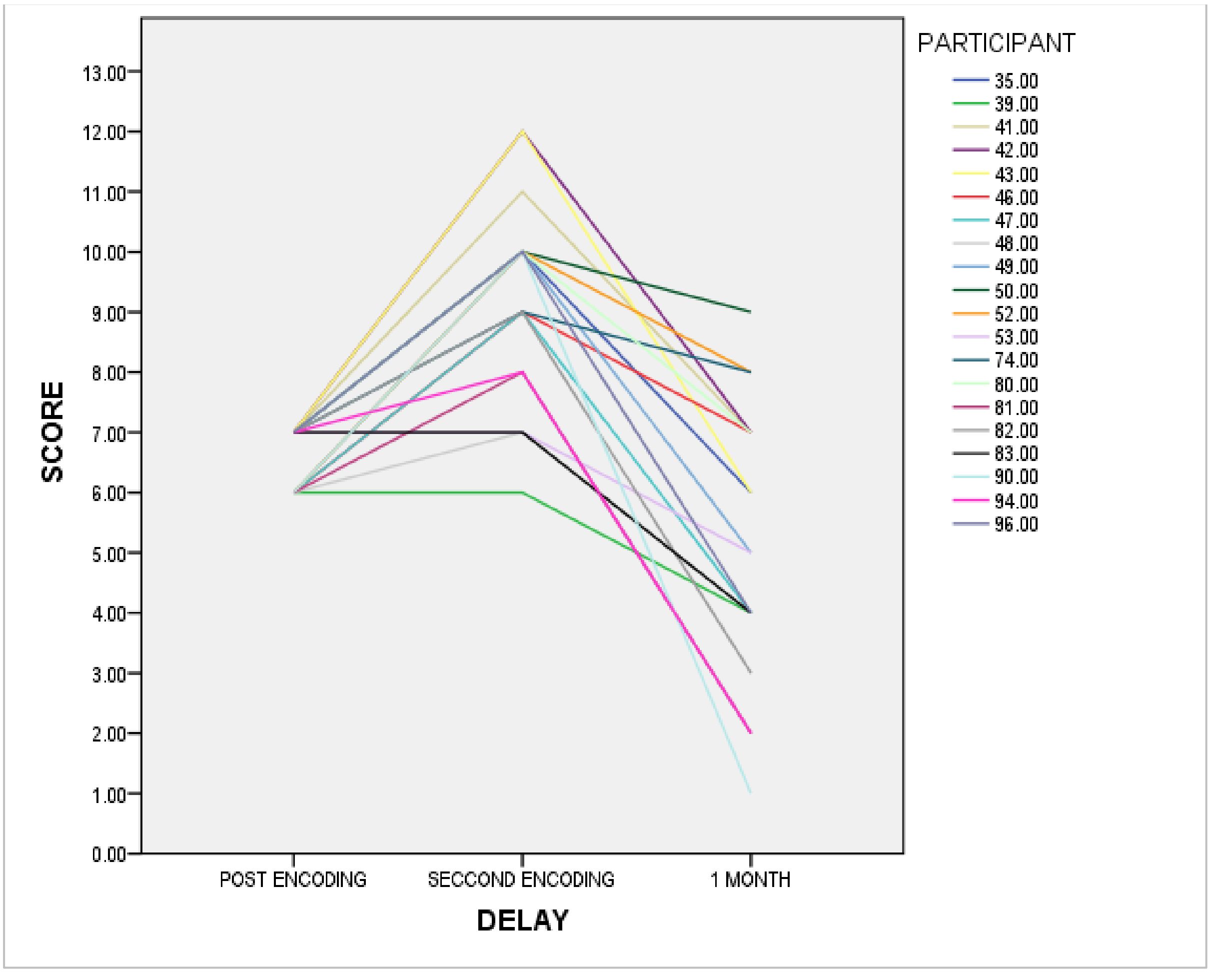


Fig. 5.4a. Individual recall performance at first post-encoding retrieval assessment, second post-encoding retrieval assessment and 1-month assessment by slow learners in the high encoding condition.

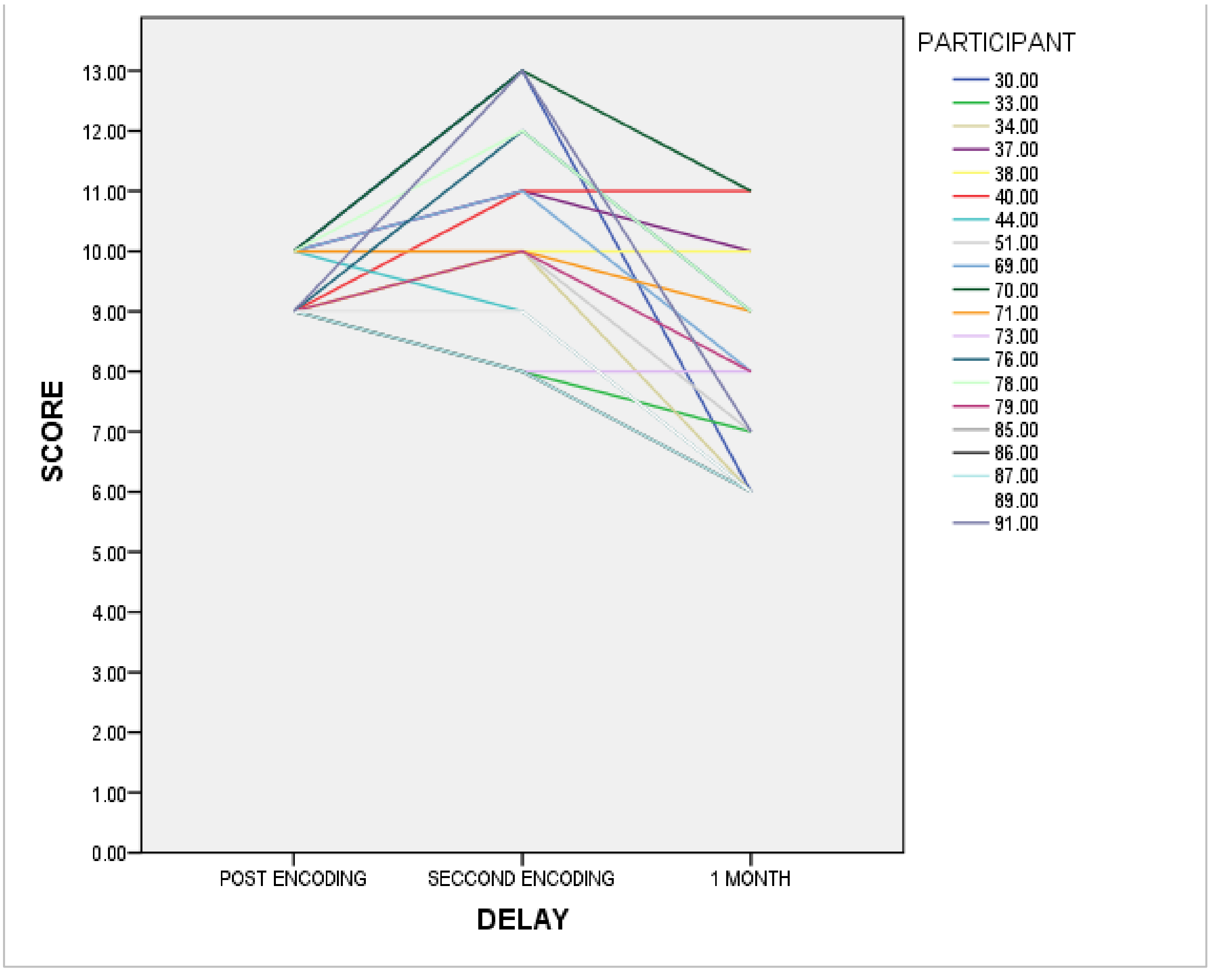


Fig. 5.4b. Individual recall performance at first post-encoding retrieval assessment, second post-encoding retrieval assessment and 1-month assessment by fast learners in the high encoding condition.

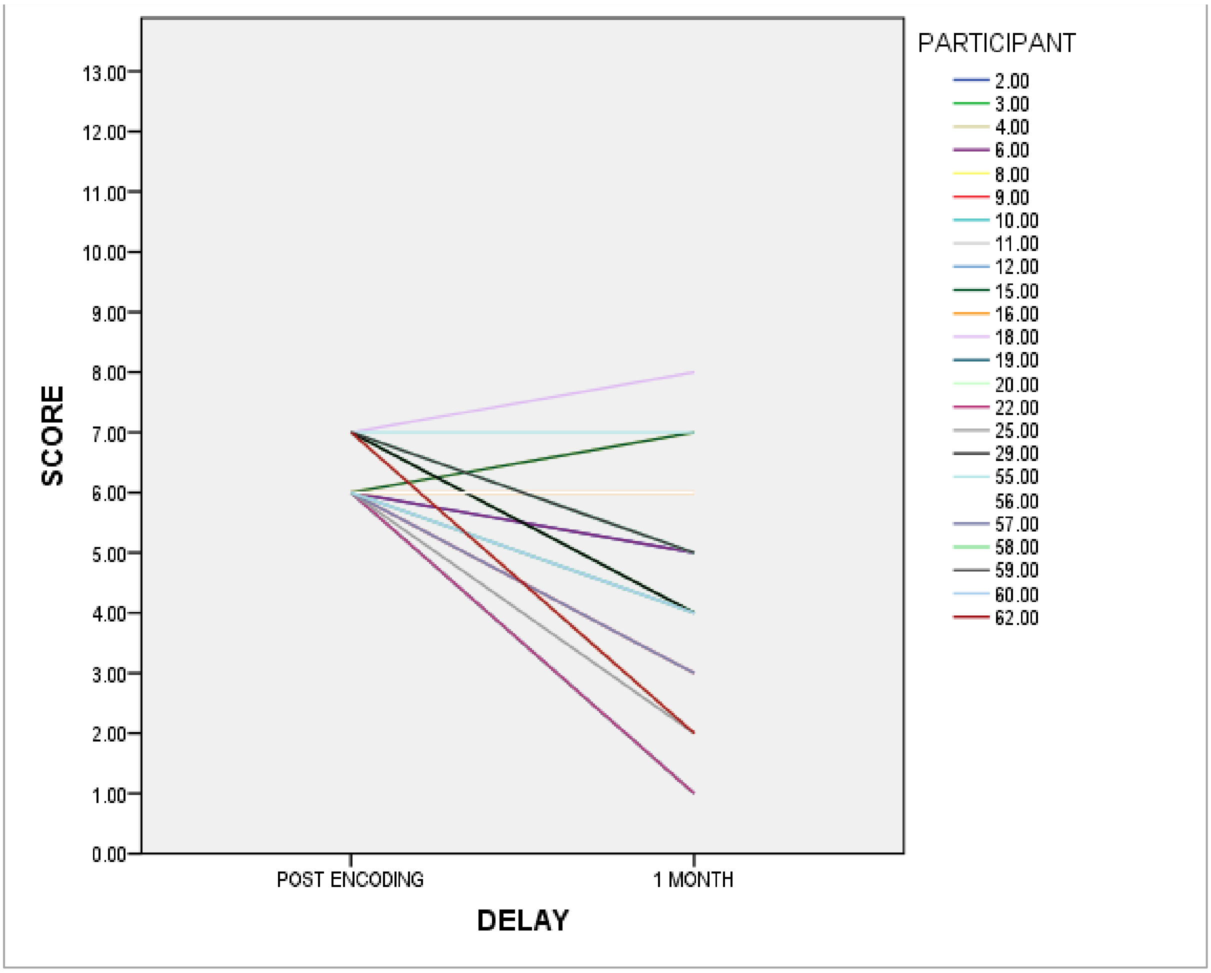


Fig. 5.4c. Individual recall performance at first post-encoding retrieval assessment and 1-month assessment by slow learners in the low encoding condition.

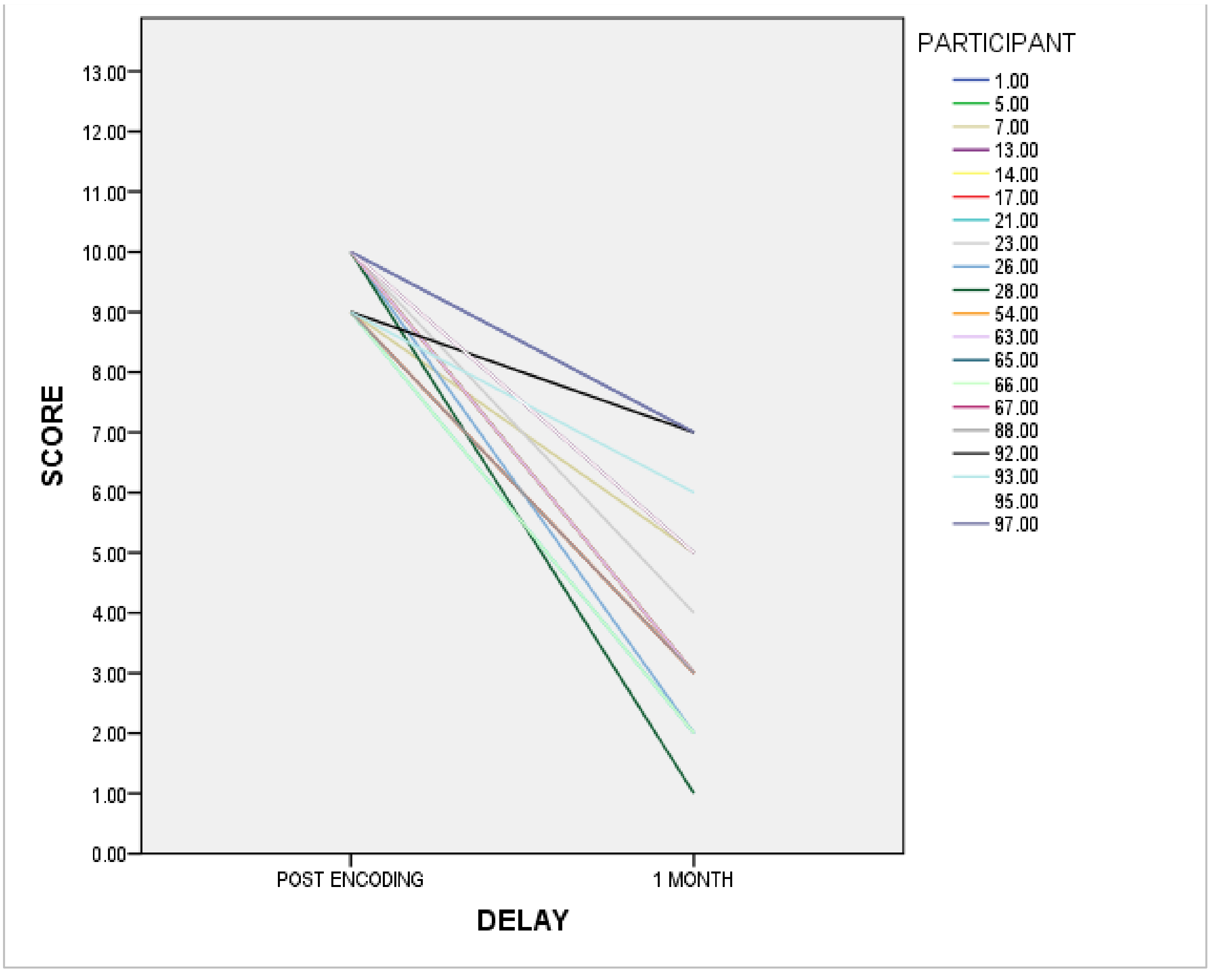


Fig. 5.4d. Individual recall performance at first post-encoding retrieval assessment and 1-month assessment by fast learners in the low encoding condition.

4. Supplementary material Experiment 7.

Experimental procedure: